

Chapter 1

Introduction

According to the National Assessment of Educational Progress (NAEP), just over 70 percent of students nationally arrive in high school with reading skills that are below “proficient” — defined as demonstrating competency over challenging subject matter. Nearly half of these students do not exhibit even partial mastery of knowledge and skills that are fundamental to proficient work at grade level.¹ These limitations in literacy skills are a major source of course failure, high school dropout, and poor performance in postsecondary education.² While research is beginning to emerge about the special needs of striving adolescent readers, very little is known about effective interventions aimed at addressing these needs.³

To help fill this gap and to provide evidence-based guidance to practitioners, the U.S. Department of Education initiated the Enhanced Reading Opportunities (ERO) study — a demonstration and rigorous evaluation of supplemental literacy programs targeted to ninth-grade students with limited literacy skills.⁴ The demonstration involves 34 high schools from 10 school districts that are implementing one of two supplemental literacy programs: Reading Apprenticeship Academic Literacy, designed by WestEd, or Xtreme Reading, designed by the Kansas University Center for Research on Literacy. These programs were selected from a pool of 17 applicants for this project by a national panel of experts on adolescent literacy. The programs are supplemental in that they consist of a year-long course that replaces a ninth-grade elective class rather than a core academic class. They aim to help striving adolescent readers develop the strategies and routines used by proficient readers and to motivate them to read more and to apply these strategies to a wide range of texts.

The evaluation is assessing the impact of the two supplemental literacy programs on students’ reading comprehension skills and on their general performance in high school, including achievement on standardized tests, course completion, and progress toward graduation. MDRC — a nonprofit, nonpartisan social policy research organization — is conducting the evaluation in partnership with the American Institutes for Research (AIR) and Survey Research Management (SRM).

¹Lutkus, Rampey, and Donahue (2006) provide an analysis of NAEP reading results for urban school districts in the context of the national NAEP performance trends.

²Carnevale (2001); Kamil (2003); Snow and Biancarosa (2003).

³Biancarosa and Snow (2004).

⁴The ERO study is known more formally as “An Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies.”

The evaluation is based on a two-level random assignment research design. In the first stage, 34 participating high schools were randomly assigned to use one of the two supplemental literacy programs. In the second stage, more than 2,900 eligible students from these high schools (students with reading test scores between two and five years below grade level) were randomly assigned either to participate in one of the literacy programs or to continue in a regular elective class.

Evaluation data were collected with a standardized reading comprehension test and a survey that were administered to students at two points during the ninth-grade year: (1) a baseline assessment and survey at the start of ninth grade and (2) a follow-up assessment and survey at the end of ninth grade. The study also includes observations of the supplemental literacy classes and interviews with teachers and administrators in each of the high schools, to learn about the fidelity of program implementation.

This report presents early findings from the ERO study, based on the first year that the supplemental literacy programs were in operation. It focuses on the first of two cohorts of ninth-grade students from each of the participating high schools. The report assesses the impact that the two supplemental literacy programs had on these students' reading comprehension skills through the end of their ninth-grade year. The report also presents impacts on selected reading behaviors, as a secondary indicator of the programs' potential effect on the initial cohort of students. The report provides an assessment of the fidelity with which the programs were implemented and discusses factors that influenced the capacity of the schools and teachers to operate them as intended over the course of the study's first year.

The early findings presented in this report should be seen as preliminary because of the implementation challenges that arose from the rushed start of the project and that are often typical of the initial phases of complex demonstrations. Also, while the end of ninth grade and the end of students' exposure to the literacy programs is a useful point at which to assess impacts on reading comprehension skills, the evaluation does not yet include information on students' longer-term performance in high school. This means that it is too early to draw definitive conclusions about the potential of these literacy interventions to improve the performance of striving adolescent readers.

In anticipation of these challenges, the U.S. Department of Education extended the demonstration and evaluation to include a second cohort of ninth-grade students who would be exposed to the programs during their second year of operation. Two subsequent reports from the ERO study will provide stronger evidence about program impacts and implementation. The second report will focus on the second year of implementation and on the second cohort of ninth-grade students to enter the study sample. In the second year of the study, most of the schools did not experience the start-up delay that they encountered in the first year. Thus, in

most of the participating schools, findings for the second cohort of students will reflect their exposure to a full year of program operation and to teachers who were more experienced in implementing the programs. The third report will focus on the longer-term impacts on students' academic achievement in tenth and eleventh grades, including their performance on high-stakes state tests and their progress toward graduation.

The remainder of this chapter describes the nature and consequences of the low literacy levels with which many students enter high school — a key motivation for the ERO study. It also provides a more detailed description of the ERO demonstration and of the research design being used to assess the impact of the two supplemental literacy programs selected for the project.

Striving Adolescent Readers: The Nature and Consequences of the Problem

The ERO study emerged from the growing recognition of the role that limited literacy skills play in restricting student success throughout high school and, particularly, during the tenuous transition from eighth to ninth grade. Some view large, comprehensive high schools as impersonal, bureaucratic, anonymous, and unable to respond effectively to the diverse needs of adolescents.⁵ Such schools can be especially inhospitable to ninth-graders — particularly to students with weak academic preparation, especially in literacy — and can exacerbate feelings of low self-efficacy and social marginalization.⁶ Further, as students progress through the primary grades to the middle grades and then to high school, they read increasingly complex textbooks, supplementary materials, and electronic text. In particular, the reading requirements of ninth grade represent a new and giant leap for entering freshmen, who face an increase in the amount of reading that is required in their courses, textbooks that are thicker and more intimidating than in previous grades, and a vocabulary load in content-area instruction that can be overwhelming. Struggling readers — who may harbor real interest in their academic subjects but lack confidence in their ability to improve their reading — may feel uncomfortable in school, may increasingly avoid challenging reading materials, and may try to avoid situations in which their poor reading skills will be exposed.⁷

Recent research indicates that struggling adolescent readers grapple with a constellation of reading difficulties that range from severe problems with basic literacy skills to troubles gaining a nuanced understanding of text. According to a report issued by the Southwest Educational

⁵National Association of Secondary School Principals (1996); Darling-Hammond, Aness, and Ort (2002); Sizer (1984); Harvey and Housman (2004).

⁶Legters and Kerr (2001); Lee, Bryk, and Smith (1993); Shanahan (2004).

⁷Guthrie (2002); Guthrie and Alvermann (1999); Wigfield (2004).

Development Laboratory, struggling adolescent readers generally demonstrate the following characteristics:⁸

1. Their reading is often slow and lacking in fluency, often because they struggle with decoding.
2. Their comprehension skills are weak, often because of limited background knowledge, difficulty making inferences, limited vocabulary, and limited self-regulation strategies.
3. They lack motivation to persist in reading.

In their report *Reading Next — A Vision for Action and Research in Middle and High School Literacy*, Biancarosa and Snow indicate that about “70 percent of older readers require some form of remediation.”⁹ However, these students’ problem is less often with knowing how to read words on a page and rather more often with understanding what they read; that is, they have difficulties with comprehension.¹⁰ Their struggles with comprehension can stem from lack of fluency (they cannot read quickly enough to facilitate comprehension) or from a lack of strategies for how to make sense of what they read or even from a lack of experience employing such strategies across a variety of types of texts in different situations. The goal for these readers is to advance from basic literacy skills to mastering the reading comprehension skills necessary for success in secondary school and beyond. That is, although some adolescent readers may still need support with basic reading skills — decoding, phonics, phonemic awareness, and so on — the majority need additional support and instruction to become expert readers who can move through complex passages containing advanced vocabulary — with fluency and the ability to derive the intended meaning.¹¹

Most high schools provide no formal instructional supports for literacy development, and most English/language arts and social studies teachers do not see literacy development as within their purview. Researchers have noted some common attitudes toward and assumptions about literacy instruction in high schools that may account for this gap. Most significantly, high school teachers view literacy skills as functional tools to be employed in the service of content-area learning.¹² Roe, Stoodt, and Burns suggest that secondary school instructional planning also reflects the belief that teaching reading is the domain of elementary schools, that teaching reading in the content areas is separate from teaching subject matter, that teaching reading in

⁸Peterson et al. (2000).

⁹Biancarosa and Snow (2004) focus on students in grades 4 through 12.

¹⁰Curtis and Chmelka (1994).

¹¹Schoenbach, Greenleaf, Cziko, and Hurwitz (1999).

¹²Bloome (2001); Dillon, O’Brien, and Volkmann (2001); O’Brien, Moje, and Stewart (2000).

secondary schools means teaching remedial reading, and that teaching reading is the purview of English teachers or reading specialists outside content classrooms.¹³ According to Shanahan, English teachers also do not assume that they should be the ones to teach struggling readers the skills they need.¹⁴ Shanahan further notes the belief of content-area teachers (including English teachers) that if they attempt to teach reading-across-the-curriculum strategies, they will only be taking valuable instructional time away from their designated subject areas.

In short, gaps in the literacy skills of striving adolescent readers and the lack of internal capacity to fill these gaps raise a critical challenge for high school reform initiatives that aim to improve low-performing high schools.¹⁵ These problems are especially acute as students navigate the transition into high school and face a variety of new challenges that can easily push them off the path toward graduation and preparation for postsecondary education and the labor market. Over the past several years, education researchers and practitioners have developed new strategies to address the challenges that ninth-grade students face as they enter high schools, but few have tackled directly the range of problems that arise from limited literacy skills.

Key Elements of a Response and the Role of Supplemental Literacy Programs

In an attempt to mitigate the difficulties that ninth-graders face as they make the transition to high school, many schools are beginning to adopt a range of targeted and comprehensive reform initiatives. Increasingly, these initiatives begin with changes in the structure and organization of the high school through the creation of “smaller learning communities” (SLCs) or even small, independent schools.¹⁶ These structural reforms are often accompanied by curricular and instructional reforms, some of which may be targeted to students who enter high school with limited literacy and math skills.¹⁷ The ERO project builds directly on this precedent by embedding supplemental literacy interventions in “Freshman Academies” — SLCs composed solely of ninth-grade students. To set the context for the ERO study, the following summarizes the roles that SLCs are increasingly playing in high school reform initiatives.

Typically, SLCs function as “schools within schools” characterized by groups of 100 to 200 students who take at least a core set of classes together from interdisciplinary teacher teams. SLCs seek to foster a personalized atmosphere in which students and teachers come to know and trust each other and hold each other to high standards. In Freshman Academies, ninth-

¹³Roe, Stoodt, and Burns (1998).

¹⁴Shanahan (2004).

¹⁵Quint (2006).

¹⁶Abrams and Oxley (2006).

¹⁷Quint (2006).

graders are grouped into a section of the high school building or in an entirely separate building, where they receive extra support from teachers, counselors, and mentors. Several studies suggest that these academies can be effective structures for supporting students as they make the difficult transition from middle school to high school. These studies indicate that SLCs for ninth-grade students can produce increases in attendance, credit accumulation, and on-time promotion to the tenth grade.¹⁸

Despite the growth of SLCs as a central component of high school improvement strategies, high school reformers have increasingly come to acknowledge that changes in instruction and academic supports may be necessary but are insufficient alone to improve the academic performance of struggling students. While formal literacy instruction is not widely practiced in specific content-area classrooms, supplemental reading programs have been developed to respond to the needs of students who have weak literacy skills. Implementing these interventions within SLCs and Freshman Academies can also provide a particularly strong, supportive structural foundation on which to implement and sustain high-quality instructional interventions, such as supplemental literacy programs. Developmental theory suggests that, from both students' and teachers' perspectives, such instructional changes may be more effective when they are mounted within settings that also attend to students' socioemotional needs.¹⁹

Recently, researchers have begun to identify elements of interventions that are designed to address the literacy needs of struggling adolescent readers. At the same time, very few of these elements have been subjected to rigorous evaluations either alone or in combination with one another. Thus, there has been a growing demand for better evidence about what works, for whom, and under what conditions.²⁰ As described below, the elements of these intervention strategies encompass content-related features and the framework for their implementation.

Content-Related Features²¹

- *Motivation and behavior.* Addresses the question, “Why read?” Includes cooperative learning environments and use of high-interest materials.

¹⁸Quint, Miller, Pastor, and Cytron (1999); Kemple and Herlihy (2004); Kemple, Connell, Legters, and Eccles (2006).

¹⁹Kemple, Connell, Legters, and Eccles (2006).

²⁰Alliance for Excellent Education (2004); Alvermann (2002); Biancarosa and Snow (2004); Guthrie and Alvermann (1999); Kamil (2003); National Reading Panel (2000); RAND Reading Study Group (2002); Snow and Biancarosa (2003).

²¹National Reading Panel (2000); Beck, McKeown, and Kucan (2002); RAND Reading Study Group (2002); Snow and Biancarosa (2003); Biancarosa and Snow (2004).

- *Advanced phonics and decoding.* Accounts for the range of expertise in adolescents’ mastery of alphabetic sounds and word decoding. Uses word study that teaches how to decode while simultaneously teaching meaning.
- *Fluency.* Uses guided oral reading at students’ individual reading levels. Includes practice with expository and narrative text.
- *Vocabulary.* Teaches strategies to identify and learn new words and to build context for new words and concepts. Uses both direct and indirect techniques for teaching vocabulary.
- *Comprehension.* Teaches components of text structure, generically and with specific reference to content-area learning. Uses both modeling and instruction to teach strategies and thought processes. Activates students’ prior knowledge and encourages higher-order thinking.
- *Metacognition.* Teaches students to reflect on how they read, to recognize faulty comprehension, and to apply “fix-up” strategies.
- *Writing.* Teaches a process for writing (planning, writing, feedback, editing) that will be successful across the high school curriculum. Promotes use of higher-order thinking skills.

Implementation Framework²²

- *Instructional approach.* Relies on both direct comprehension instruction and student self-directed learning. Includes whole-group, small-group, and individualized instruction. Instruction should be embedded in content and should link concepts, skills, and strategies across topics and over time.
- *Scheduling and duration.* Provides students a minimum of 225 minutes of literacy instruction per week (organized as 45-minute classes each day or as 80- to 90-minute blocked classes every other day), over and above the regular English or language arts classes. Includes lessons or instructional segments that can extend for a full academic year.
- *Group size.* Can accommodate up to 15 students per period to facilitate multiple modes of instruction and attention to individual needs.

²²Biancarosa and Snow (2004).

- *Materials.* Includes diverse reading materials, highly engaging and appropriate for age and skill level.
- *Use of technology.* Uses technology for practice of skills and strategies presented by the teacher.
- *Teacher training and support.* Includes intensive introductory training followed by on-site coaching and ongoing technical assistance. Provides teachers with resources and guides to conduct instruction and assess student progress.
- *Assessment.* Includes regular assessment of reading skills and ties the results to instruction. Uses assessment both to diagnose problems and to monitor progress.
- *Cost.* Must be affordable to allow for adoption by low-income districts.

An array of programs has been developed with one or two of these elements embedded in them.²³ Yet, very little has been done to develop an overall strategy for directing and coordinating a multidimensional response to the needs of students who face the greatest risk of school failure by virtue of their limited literacy skills. In their high-profile call to action to address the needs of struggling adolescent readers, Biancarosa and Snow call for a series of demonstrations that attend to the challenges and variations associated with different components, implementation strategies, and contexts and that are subject to a rigorous assessment of their impact on participating students.²⁴ The ERO study represents a direct and systematic response to this call to action.

Overview of the ERO Study

The ERO study is both a demonstration of two supplemental literacy interventions across a range of contexts and a rigorous evaluation of the interventions' impact on students' reading comprehension skills and their academic performance as they move through high school. The study is a collaboration between policy and research interests that encompass practical responses to important educational problems and a commitment to learning whether these responses produce their desired effects. The U.S. Department of Education's Office of Elementary and Secondary Education (OESE) is providing direct support for implementation to the participating schools and districts, while its Institute of Education Sciences is overseeing the design and execution of the evaluation effort. Incorporating the evaluation expertise of the re-

²³For a summary of the evidence base on interventions that incorporate the elements listed above, see Biancarosa and Snow (2004).

²⁴Biancarosa and Snow (2004), p. 23.

search team, the substantive knowledge of the model developers and the operational capacity of participating sites, the ERO project places a useful policy instrument at the service of both helping students and building knowledge. Following is a brief overview of the demonstration and evaluation components of the ERO study.

A Demonstration of Supplemental Literacy Interventions

The ERO study tracks the implementation of two established supplemental literacy interventions that were developed for high school students whose reading skills are two or more years below grade level as they enter high school. Both programs incorporate many of the design elements discussed above including careful attention to student motivation, a focus on reading fluency, vocabulary, and comprehension, development of metacognition to promote reflective reading strategies, and use of technology. Each program is a full-year course that substitutes for a ninth-grade elective class and is scheduled for a minimum of 225 minutes of instruction per week. They are both designed to accommodate class sizes of 12 to 15 students.

As part of their proposal to participate in the ERO study, the developers of both programs provided suggestive evidence of their developmental appropriateness for the target population of students and of their alignment with the available research base on strategies for improving the literacy skills of struggling adolescent readers.²⁵ Each intervention was part of a larger and more comprehensive high school reform initiative. For the purposes of the ERO study, the programs were modified somewhat and adapted for implementation as an independent class that would replace a regular elective class for ninth-grade students. In order to meet the special needs of high school teachers who do not have reading instruction credentials, the programs' developers also intensified their professional development and coaching strategies. While the two programs share core goals and many instructional strategies, they differ primarily in their approach to implementation.

The supplemental literacy programs are being implemented in 34 high schools from 10 school districts across the country. The districts were selected through a special grant competition organized by the U.S. Department of Education's Office of Vocational and Adult Education (OVAE).²⁶ Experienced, full-time English/language arts or social studies teachers volun-

²⁵For an overview of research related to Reading Apprenticeship Academic Literacy, see Schoenbach, Greenleaf, Cziko, and Hurwitz (1999). For an overview of research related to Xtreme Reading and the Strategic Instruction Model, see Schumaker and Deschler (2003, 2004).

²⁶For a complete application package for the special competition, see U.S. Department of Education (2005). The special grant competition was part of OVAE's Smaller Learning Communities initiative and was designed to provide extra funding to qualifying districts for the implementation of the supplemental literacy programs and participation in the ERO evaluation. The grants also included funds for general support of the Smaller Learning Communities initiatives under way in the districts. In 2006, responsibility for the Smaller Learning Communities initiative and for the special ERO grants was moved from OVAE to OESE.

teered to teach the programs for a period of two years. It should be noted that the participating sites were not selected to be representative of all districts and schools across the country. As a result, findings from the ERO study cannot be generalized statistically to the full population of districts and high schools or to urban districts and schools. At the same time, the participating sites reflect much of the diversity of midsize and large urban school districts that serve low-income and disadvantaged populations of students. Thus, the findings will be widely applicable and highly relevant to districts and high schools that are struggling to meet the needs of ninth-graders who lack the literacy skills required for academic success.

A Rigorous Impact Evaluation

The ERO evaluation will unfold over a five-year period and will address the following questions:

- What are the short-term impacts of these supplemental literacy interventions on ninth-grade students' reading skills and behaviors?
- For which subgroups of students are supplemental literacy interventions most or least effective?
- What factors promote or impede successful implementation of the supplemental literacy interventions? In what ways are implementation fidelity and quality associated with program impacts (or lack of impacts) on reading achievement and other outcomes?
- What are the longer-term impacts on other academic outcomes, such as achievement on high-stakes standards-based assessments, performance in academic courses, and progress toward graduation? What is the nature of the relationship between the impacts on reading skills and the impacts on these other outcomes?

The current report provides an early assessment of the first three of these questions as reflected in the first year of implementation. Subsequent reports will provide evidence about the effectiveness of maturing versions of the programs and will address the questions about longer-term impacts.

The ERO evaluation utilizes a two-level random assignment research design. First, within each district, eligible high schools were randomly assigned to use one of the two supplemental literacy programs. This feature of the design allows a direct comparison of the effectiveness of the two programs and avoids confounding the effect of purposeful or self-selection of schools to use the two programs with a true difference in the programs' impact on student achievement.

The second feature of the study design involves the random assignment of eligible and appropriate students within each of the participating high schools. Each high school was asked to identify at least 100 ninth-grade students who were reading at least two years below grade level. Approximately 55 percent of these students were randomly assigned to enroll in the ERO class, and the remaining students make up the study's control group and enrolled in or continued in a regularly scheduled elective class. This feature of the design is possible because there were more eligible and appropriate students in each high school than the 50 to 60 students that the literacy programs are able to serve. Students in both groups take the regular English/language arts classes offered by their schools as well as other core academic and elective classes required of or offered to ninth-graders. The study includes two cohorts of ninth-grade students: one cohort that was enrolled in the study at the beginning of the 2005-2006 school year and one cohort that was enrolled in the study starting in the 2006-2007 school year.

Finally, the ERO evaluation taps a variety of data sources to measure students' reading achievement and school performance and to assess the fidelity of program implementation.

Overview of This Report

The remaining chapters in this report provide further background on the study design and discuss the implementation and impact findings. Chapter 2 describes the sample of schools and the first cohorts of students who are participating in the study. Chapter 3 presents an in-depth description of the two supplemental literacy programs and their implementation during the initial year of the study. Chapter 4 examines student enrollment and attendance in the ERO classes and looks at the rate at which students in the study's non-ERO sample participated in supplemental literacy services both in and outside school. Chapter 5 reports on the early impacts of the literacy interventions.

This report provides an early look at the implementation and impact of the two literacy interventions based on their initial year of operation in the participating schools. Because of the late award of the special SLC grants, none of the high schools was able to begin its program at the start of the school year. Also, the schools and teachers had no prior experience with the programs, and their knowledge and expertise evolved throughout the year. The delay in program start-up and the schools' and teachers' evolving competence with them means that the interventions did not receive as complete a test as would be expected with a full year of operation and prior experience with implementation. As a result, the findings presented in this report should be interpreted cautiously in terms of their implications for education policy and practice. Later reports on the ERO evaluation will provide more conclusive evidence about the effectiveness of these interventions and a more solid footing for use by policymakers and practitioners. Despite the limitations of an early assessment of program experiences, the current report aims to offer useful insights into the characteristics, implementation, and impact trends of these interventions.

Chapter 2

Study Sample and Design

This chapter describes the sample of schools and students involved in the Enhanced Reading Opportunities (ERO) study, the different sources of data and the impact measures created from these data, student response rates during follow-up data collection, and the analytic methods used to assess program impacts. The chapter discusses the following key points:

- Thirty-four schools from 10 school districts were selected for the study and were randomly assigned to use one of the two supplemental literacy programs. The resulting two groups were similar on a range of school characteristics.
- The study sample includes 2,916 students with baseline reading test scores that fell between two and five years below grade level. Fifty-seven percent of these students were randomly assigned to the ERO group and were scheduled into the ERO classes, and the remaining 43 percent were assigned to a non-ERO control group and continued in a ninth-grade elective class.
- Approximately 83 percent of the students in the study sample (a total of 2,413 students) completed the follow-up reading assessment and survey. Among respondents, overall differences found in background characteristics between the ERO and non-ERO groups are not statistically significant.
- Statistical-power calculations indicate that the full study sample available for the impact analysis is sufficient for minimum detectable effects sizes of 0.06 standard deviation units or larger for the reading test score outcomes. The samples available for each of the two supplemental literacy programs are sufficient for minimum detectable effects sizes of 0.10 standard deviation units or larger.

School Sample

The school districts participating in this study were selected through a special grant competition run by the Office of Vocational and Adult Education (OVAE) within the U.S. Department of Education (ED).¹ As an extension of the Smaller Learning Communities (SLCs) grant program, this competition sought to provide funding for the implementation of two sup-

¹U.S. Department of Education (2005).

plemental ninth-grade literacy programs in selected high schools and to sustain and enhance existing SLCs in these high schools.

In June 2005, ED selected 10 grantee school districts encompassing 34 high schools — from a pool of 33 applicant districts.² The 10 grantee districts encompass 65 high schools, with the smallest district having four high schools and the largest having 22 high schools. Seven of the grantee districts included four of their high schools in the study, and the remaining three districts included two high schools. Grantee districts will receive approximately \$1.25 million over five years for each participating high school. From their SLC grants, districts were required to set aside \$250,000 per high school over the first two years of their grant period to cover the costs of implementing the supplemental reading programs, including costs associated with teachers' salaries and benefits, teacher-training activities, coaching and materials to be provided by the program developers, classroom computers, and other equipment and materials.

Random Assignment of Schools

Following the selection of grantee districts to participate in the ERO study, the study team randomly assigned the participating schools to implement one of the two literacy programs. Within each district, half the participating schools were randomly assigned to the Reading Apprenticeship Academic Literacy program, and half were randomly assigned to Xtreme Reading. Schools were randomly assigned to the interventions as a safeguard against selection bias. That is, if districts and developers had been allowed to choose the allocation of the interventions, the potential would have existed for decisions to have been made based on any of a variety of characteristics associated with outcomes of overall effectiveness that might have made one school a more favorable candidate over another for a more “successful” implementation of the program. Such characteristics cannot be measured, thereby presenting a possible threat to the validity of the study. Essentially, by randomly assigning schools to one of the two supplemental literacy interventions, the study ensured that the intervention developers could not select schools that were higher performing or at a higher level of readiness for their programs. It also ensured that the schools could not select a literacy program that they believed would be more appropriate or more effective for their school. As a result, differences in impacts that may emerge between the two groups of schools can be attributed to differences between the two programs rather than to differences in school characteristics or the method for assigning schools to the programs.

²The number of applicants for the special SLC Grant Competition was reported to the study team by OVAE staff.

Characteristics of Schools Selected for the ERO Project

Table 2.1 presents characteristics of the 34 high schools participating in the ERO study. Overall, ERO programs were implemented in schools located predominantly in large and mid-size cities, with some of the schools in each of these categories being listed as “urban fringe.” As specified by the OVAE grant requirements, all schools enrolled more than 1,000 students in grades 9 through 12, averaging 1,685 students per school. The schools enrolled an average of 570 ninth-grade students, ranging from 320 to 939 ninth-grade students per school. Table 2.1 shows the average “promoting power” for the participating schools, which can serve as a proxy for the likely longitudinal graduation rate.³ It indicates that the twelfth-grade class is 59 percent of the size of the ninth-grade class three years earlier, suggesting that roughly 41 percent of students have left the schools between the ninth and twelfth grades. The table also shows that 38 percent of the students in the participating schools were eligible for Title I services and that 47 percent of the students were approved for free or reduced-price lunch.

Overall, Table 2.1 indicates that there is a high degree of similarity between the schools randomly assigned to use Reading Apprenticeship Academic Literacy and the schools assigned to use Xtreme Reading.

Table 2.1 also includes information about all high schools across the country that, like those selected for the ERO study, are located in large and midsize cities, served over 1,000 students in grades 9 through 12, and did not select students based on past achievement or performance. This national census of similarly situated high schools provides a reference point that helps contextualize and describe the ERO high schools. In comparison with the national sample, the schools selected for the ERO study include a higher proportion of students with characteristics associated with low performance. The ERO schools have lower levels of student promotion, higher percentages of students eligible for free and reduced-price lunch, and higher eligibility for Title I funding. Additionally, the populations at ERO schools comprise higher percentages of minority students than the national sample.

Student Sample

At the inception of the ERO project, the primary target population for the supplemental literacy interventions included students entering ninth grade with reading skills that were between two and four years below grade level. To qualify for an ERO grant, districts were re-

³Balfanz and Legters (2004) developed this measure of “promoting power” to approximate a school’s graduation rate. It is calculated as the ratio of the number of twelfth-grade students in a given school year to the number of ninth-grade students from three years prior.

The Enhanced Reading Opportunities Study

Table 2.1

**Characteristics of ERO Schools and Average Schools
in the United States (2004-2005)**

Characteristic	All ERO Schools	Reading Apprenticeship Schools	Xtreme Reading Schools	Average U.S. Schools ^a
Average number of students	1,685	1,687	1,683	1,866
Average number of students in grade 9	570	566	574	556
Average number of students in grade 10	432	436	429	478
Average number of students in grade 11	358	359	358	424
Average number of students in grade 12	317	312	322	382
Average promoting power ^b (%)	59.1	56.7	61.6	75.4
Students eligible for free or reduced-price lunch (%)	46.9	44.5	49.2	30.0
Race/ethnicity (%)				
Hispanic	25.1	24.6	25.6	19.3
Black	41.1	41.9	40.4	19.7
White	31.2	31.0	31.5	53.5
Other	2.6	2.6	2.6	7.0
Eligible for Title I (%)	38.2	41.2	35.3	26.0
Locale (%)				
Large city ^c	52.9	52.9	52.9	61.2
Midsize city ^d	47.1	47.1	47.1	38.8
Sample size	34	17	17	3,727

SOURCES: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey Data", 2004-2005 and 2001-2002.

NOTES: This table provides information on 34 ERO schools from 10 districts.

Rounding may cause slight discrepancies in calculating sums and differences.

^a"Average U.S. Schools" includes schools that have more than 1,000 total students, have more than 100 students in each grade during 2004-2005, have at least 125 students in the ninth grade during 2001-2002, are noncharter schools, are located in a large or midsize city or in the urban fringe of a large or midsize city, are defined as "regular" schools by the Common Core of Data, and are operational at the time of the Common Core of Data report.

^b"Promoting power" is calculated as the ratio of twelfth-grade students in 2004-2005 to ninth-grade students in 2001-2002.

^c"Large city" is defined as a city having a population greater than or equal to 250,000. Schools in this category also include the urban fringe of a large city.

^d"Midsize" city is defined as a city having a population less than 250,000 but greater than 50,000. Schools in this category also include the urban fringe of a midsize city.

quired to provide documentation that each high school would include at least 125 ninth-grade students with reading skills at these levels.⁴

Among the first tasks for the ERO study were to identify potentially eligible students in each of the participating high schools, obtain parental consent for the students to be included in the study sample, and administer a baseline reading test and a survey. Then, assuming that 125 students were eligible for the ERO programs and consented to be in the study, the study team would conduct random assignment such that up to 60 of these eligible students would be selected to enroll in the ERO classes. Of the students randomly assigned to the ERO program, the school was responsible for scheduling those students into four ERO class sections. Typically, those sections each contained 12 to 15 students. Of the remaining 65 students, up to 50 students would be assigned to enroll or remain in a regular ninth-grade elective class. The remaining 15 students would constitute a nonresearch waiting list and would be admitted to an ERO class if enrollment levels fell below the desired minimum of 12 students, due to attrition over the school year.⁵ Because the special SLC grants were not awarded until the summer of 2005, this process could not begin until the start of the 2005-2006 school year. This meant that the student study sample would not be identified until several weeks into the school year and that students selected for the ERO classes would be forced to withdraw from an elective course they had already begun to attend.

Early in the 2005-2006 school year, it became clear that the study team and the schools were facing significant challenges that would require some modification in the original targeting criteria and that would delay the start of the classes further. The study team was in regular contact, both in person and by telephone, with staff in the participating schools and districts to monitor the student testing and recruitment process. The team learned that several of the schools had fewer than the prescribed number of students in the target range — at least according to the reading test that was being used for the ERO study. Also, all the schools faced severe challenges in getting eligible students to return signed consent forms. As a result, the study sample was expanded to include students between two and five years below grade level, and the eligibility criteria to be in an ERO class were expanded to include students with reading levels between one and five years below grade level. Schools also employed more intensive strategies to obtain consent forms. In the end, all the participating schools were able to meet minimal targets for the study sample, but this was not completed until an average of six weeks into the school year.

⁴It should be noted that English Language Learning (ELL) and special education students who required specific classroom, instructional, or testing accommodations were not eligible for the ERO classes. The ERO programs were not designed to accommodate the special needs of these students nor the potential scheduling conflicts with other services that the students were likely to receive.

⁵Note that students assigned to the nonresearch waiting list were not included in the analysis, even if they were later scheduled into ERO classes.

Following a more detailed discussion of the student recruitment and random assignment process, this section of the chapter describes the characteristics of the core sample of students in the study's first cohort.

Student-Level Random Assignment

Because the special SLC grants were not awarded until the summer of 2005, the student recruitment process did not begin until the start of the 2005-2006 school year. Staff from each of the 34 high schools administered the Group Reading Assessment and Diagnostic Examination (GRADE) to their ninth-grade students. Students who scored between the fourth- and eighth-grade level on the GRADE reading comprehension subtests were considered eligible for the ERO classes. Eligible students were then asked to return a parental consent form giving permission to participate in the study and to enroll in the ERO classes if they were selected. Once eligible students returned a signed affirmative consent form and completed the baseline survey, they were entered into MDRC's random assignment database. While the recruitment of eligible students required the assistance of school and district staff members in communicating with parents and students and collecting consent forms, computerized random assignment of students was conducted solely by MDRC staff.

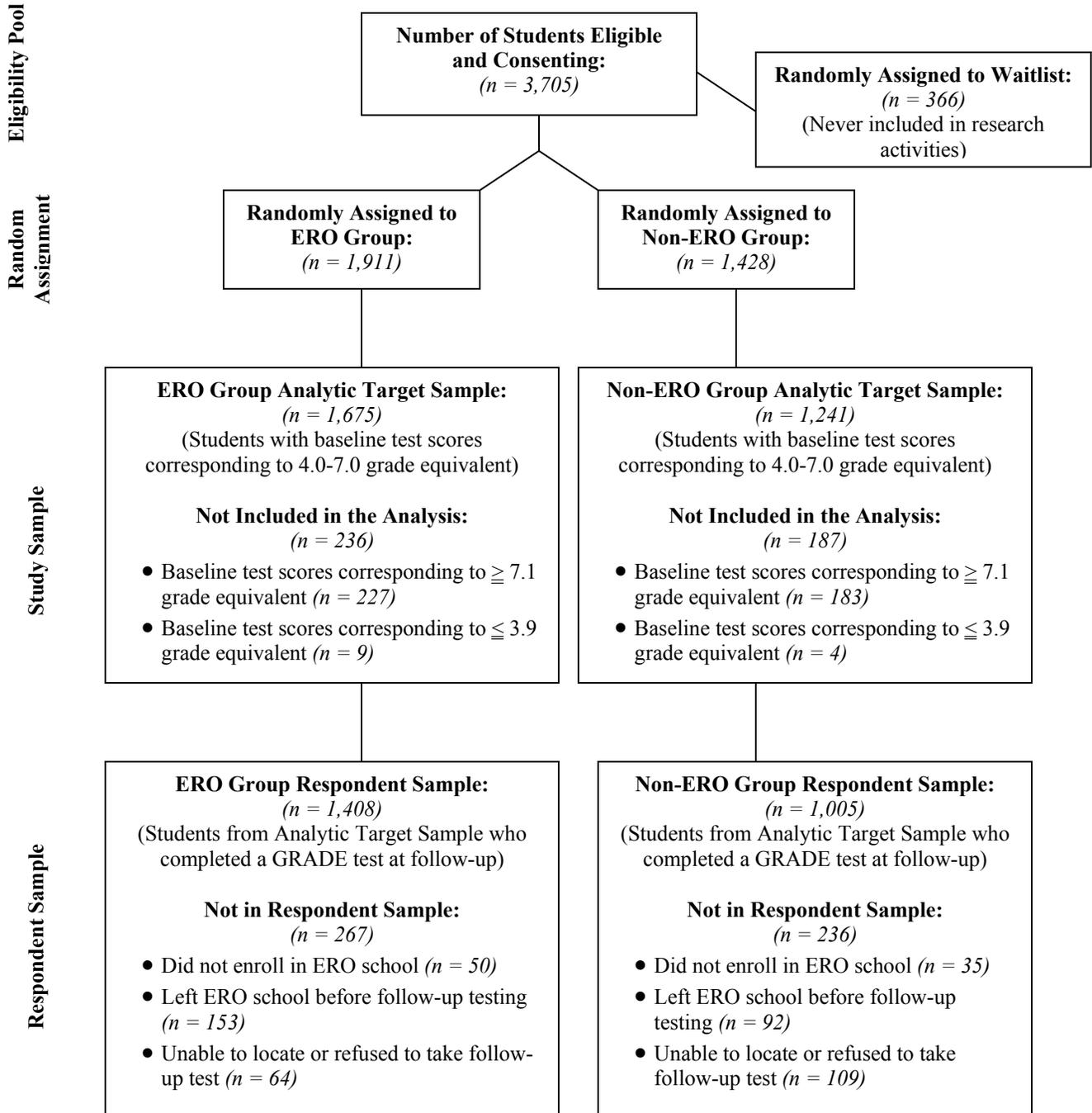
The ERO programs were designed to accommodate between 12 and 15 students per class, and each high school was required to offer four ERO class sections. The study team identified 3,339 eligible and consenting students from across the 34 participating high schools (on average, 98 students per school). Figure 2.1 shows that 1,911 (57 percent) of these students were randomly selected to enroll in the ERO classes (referred to as the "ERO group") and 1,428 (43 percent) were randomly assigned to the control group (referred to as the "non-ERO group"). Although the eligibility criteria were expanded to include students with test scores ranging from the 4.0 to 8.0 grade equivalent to keep the classes at capacity, the analyses in this report focus exclusively on the students whose baseline test scores ranged from the 4.0 to 7.0 grade equivalent (two to five years below the ninth-grade level). Figure 2.1 shows that there are 2,916 students in this group (87 percent of the entire study sample; on average, 86 students per school), with 1,675 (57 percent) randomly assigned to the ERO group and 1,241 (43 percent) randomly assigned to the non-ERO group.⁶ All further references in this report to the "study sample" refer to students with scores ranging from the 4.0 to 7.0 grade equivalent.

⁶A total of 410 students had scores that were equivalent to the 7.1 grade equivalent or higher. In addition, 13 students had scores that were equivalent to the 3.9 grade equivalent or lower. Given that the two interventions and the evaluation were designed primarily to test the effects of supplementary literacy interventions on ninth-grade students with reading comprehension skills between the fourth- and seventh-grade levels, data for these 423 students are not included in the impact analysis for this report.

The Enhanced Reading Opportunities Study

Figure 2.1

Construction of the Impact Sample from the Eligibility Pool



Characteristics of the Study Sample

The background characteristics of the ERO group and the non-ERO group were compared to determine whether random assignment resulted in two equivalent groups. There is a high degree of similarity between the two groups' baseline characteristics, as illustrated in Table 2.2. On average, students in the study sample had a reading comprehension composite score of just under 86 standard score points on the GRADE reading assessment. This average corresponds to the 5.1 grade level (an average of almost four years below grade level at the beginning of ninth grade) and to the 16th percentile nationally. The study sample is over 70 percent Hispanic or black; about 45 percent of the students speak a language other than English at home; and about 30 percent are overage for grade (15 years old or older at the start of ninth grade, suggesting that they were retained in a prior year).⁷ A general F-test indicates that, overall, there are no systematic differences in these characteristics between the ERO and non-ERO groups in the study sample. The lack of systematic differences indicates that random assignment was successful in creating two equivalent research groups at baseline. Similar results were found when examining the background characteristics of study-sample students from the Reading Apprenticeship sites and the Xtreme Reading sites, separately.⁸

Data Sources and Measures

The ERO evaluation utilizes a variety of data sources to measure students' reading achievement and reading behaviors and to assess the fidelity and quality of program implementation. Following is an overview of the data sources utilized in the current report.

Group Reading Assessment and Diagnostic Examination (GRADE)

The GRADE is a norm-referenced, research-based reading assessment that can be administered to groups. It is meant to be a diagnostic tool to assess what reading skills individuals have and what skills need to be taught.⁹ It is used widely to measure performance and track growth of an individual student and groups of students from fall to spring and from year to year. The GRADE contains multiple subtests, including two reading comprehension subtests (sentence comprehension and passage comprehension), a listening comprehension subtest, and a vocabulary subtest. For the ERO study, the two reading comprehension subtests (Level H, Form A) were administered to all students prior to random assignment. Near the end of their ninth-grade year, students completed the two reading comprehension subtests (Level H, Form B) as

⁷National Center for Education Statistics (1990).

⁸See Appendix B.

⁹See American Guidance Service (2001a, 2001b) for technical information about the GRADE.

The Enhanced Reading Opportunities Study

Table 2.2

**Characteristics of Students in Cohort 1
Full Study Sample**

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	31.8	31.7	0.1	0.943
Black, non-Hispanic	44.6	45.4	-0.8	0.556
White, non-Hispanic	17.7	17.0	0.7	0.585
Other	5.9	5.8	0.1	0.940
Gender (%)				
Male	49.9	50.1	-0.3	0.878
Female	50.1	49.9	0.3	0.878
Average age (years)	14.8	14.8	0.0	0.152
Overage for grade ^a (%)	31.5	28.3	3.1	0.054
Language other than English spoken at home (%)	45.6	45.5	0.1	0.974
Language spoken at home missing (%)	6.7	6.8	-0.1	0.921
Mother's education level (%)				
Did not finish high school	18.1	19.0	-0.8	0.554
High school diploma or GED certificate	25.0	24.8	0.1	0.942
Completed some postsecondary education	29.3	30.2	-0.9	0.581
Don't know	20.2	18.8	1.3	0.360
Missing	7.4	7.1	0.3	0.728
Father's education level (%)				
Did not finish high school	16.8	17.9	-1.1	0.444
High school diploma or GED certificate	23.0	23.2	-0.2	0.899
Completed some postsecondary education	18.3	20.6	-2.4	0.104
Don't know	33.6	29.8	3.8 *	0.027
Missing	8.3	8.5	-0.2	0.825
GRADE reading comprehension^b				
Average standard score	85.7	86.1	-0.3	0.093
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	33.2	35.8	-2.6	0.140
5.0 - 5.9 grade equivalent (%)	29.6	27.6	1.9	0.251
4.0 - 4.9 grade equivalent (%)	37.3	36.6	0.7	0.695
Sample size	1,675	1,241		

(continued)

Table 2.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

well as the vocabulary subtest. In addition to the raw score (the total number of items answered correctly), the GRADE also provides standardized scale, normal curve equivalent, grade equivalent, percentile, and stanine scores.

The primary measure of reading achievement for this study is students' scores on the GRADE reading comprehension assessment. This component of the GRADE includes subtests that measure sentence comprehension and passage comprehension. According to the GRADE technical manual, "the purpose of sentence comprehension is to identify if the student can comprehend a sentence as a whole thought or unit."¹⁰ The GRADE technical manual also characterizes passage comprehension as measuring a student's skills in understanding an extended passage consisting of a single paragraph or multiple paragraphs.¹¹ A central objective of each of the two ERO programs is to provide students with immediate and intensive instruction in the use of strategies and skills that expert readers use to understand written texts. Thus, for the purposes of the ERO evaluation, the GRADE reading comprehension assessment serves as the primary early indicator of the programs' effectiveness.

A secondary measure of students' reading achievement is their scores on the GRADE vocabulary assessment. According to the GRADE technical manual, the vocabulary subtest is intended to measure a student's knowledge of word meanings with minimal contextual clues.¹² Each of the two ERO programs provides some instruction aimed at helping students break down word meanings through advanced decoding skills and strategies for recognizing word structures (root words, prefixes, and suffixes). Thus, the GRADE vocabulary assessment can

¹⁰American Guidance Service (2001a), p. 39.

¹¹American Guidance Service (2001a), p. 45.

¹²American Guidance Service (2001a), p. 45.

provide indication of whether these approaches increase the stock of words that students know. However, because the two ERO programs focus primarily on helping students use contextual clues to understand the meaning of words, the vocabulary subtest is seen as a secondary indicator of the programs' effectiveness.

The GRADE reading comprehension and vocabulary performance levels and impacts for the ERO and non-ERO groups are presented in standard score units provided by the American Guidance Service, which publishes the GRADE.¹³ Standard scores are a more accurate representation of a student's level of performance than raw scores because they have uniform meaning from one test period to another and from one grade level to another. Standard scores indicate how far a student's performance on the test is from the average for all students at a given grade level, and standard scores take into account the variability of scores among a nationally representative group of students in that grade. Also, standard scores on the GRADE can be compared with standard scores on other tests of reading comprehension and vocabulary.

To help the reader interpret the standard score values, impact tables also present the national grade equivalent and national percentile that correspond most closely to the average standard score for the ERO and non-ERO groups, respectively. A grade equivalent score is the grade at which a particular raw score or standard score represents the median for the test's norming population. For example, a grade equivalent score of 9.0 refers to a median performance at the beginning of ninth grade, and a 9.8 grade equivalent indicates a median performance at the end of ninth grade.¹⁴

The reading comprehension and vocabulary test score impact estimates are presented both in standard score units and in effect-size units. Effect sizes provide an indication of the magnitude of the impact estimates relative to the overall variation in test scores for students in

¹³Specifically, each student's raw scores on the GRADE subtests and composite scores were converted to standard scores based on national norms for Level H, Grade 9, Spring Testing (American Guidance Service, 2001b, pp. 30-33). Based on these norms, a standard score of 100 on the GRADE reading comprehension or vocabulary test is average for a representative group of students at the end of their ninth-grade year. The standard deviation of the standard score for both tests is 15. A standard score of 85 corresponds, approximately, to the 4.9 grade equivalent.

¹⁴Note that grade equivalents and percentiles are not equal-interval scales of measurement. Grade equivalents indicate a student's place along a growth continuum, which may not increase at regular intervals. For example, the difference between a vocabulary grade equivalent of 1.0 and 2.0 represents a greater difference in vocabulary knowledge than the difference between a grade equivalent of 8.0 and 9.0. Percentiles indicate the percentage of students in the test's norming group who performed at or below a given student's score. As such, percentiles provide information only about the rank order of students' scores; they do not provide any information about students' actual performance. Because they do not reflect equal intervals between units of measure, neither grade equivalents nor percentiles can be manipulated arithmetically. (See American Guidance Service, 2001a, pp. 55-60.) Thus, readers should exercise caution when interpreting differences in grade equivalents or percentiles between the ERO and non-ERO groups and between the baseline and follow-up tests.

the study sample. For the purposes of the impact analysis, effect sizes are calculated as a proportion of the standard deviation of the test scores for students in the non-ERO group at the end of ninth grade.¹⁵ The standard deviation for the non-ERO group reflects the expected variability in test scores that one would find in the absence of the ERO programs. The impact effect size, therefore, provides an indication of how much the ERO programs moved students along this variability in expected performance.

Student Surveys

Students in the study sample completed the Enhanced Reading Opportunities Student Baseline Survey prior to random assignment. The baseline survey includes the following information for students in the study sample: gender, race/ethnicity, age, and current high school. These data items were required for random assignment and are available for all students in the study sample. The baseline survey also includes additional background information and information about students' reading behaviors and attitudes.

The study team administered the Enhanced Reading Opportunities Student Follow-Up Survey to students in the study sample at the same time as the follow-up GRADE assessment. The impact analysis presented in Chapter 5 focuses on three measures of students' reading behavior that were derived from the survey.¹⁶

Each of the ERO programs aims explicitly to increase the amount of time that students spend reading, both for school and for their own enrichment outside school. The programs do this directly by assigning students reading activities during class and for homework. They also attempt to build students' reading skills, confidence, and enjoyment, in the hope that they will take the initiative to read more frequently and for longer periods of time on their own. The first two measures in the reading behaviors impact analysis focus on how often students read various types of texts for school and outside school. Though self-reported by students, these outcomes provide a direct indication of whether the ERO programs are increasing the amount of time that students spend reading.

Amount of School-Related Reading

This measure was constructed to reflect the self-reported number of times during the prior month that a student read each of seven different types of text for school — in school or for homework: history, science, or math textbooks; literary texts; research or technical reports;

¹⁵The standard deviation of the reading comprehension standard score for the non-ERO group at follow-up is 10.458. The standard deviation of the vocabulary standard score for the non-ERO group is 10.505.

¹⁶A list of the survey items used to create these three measures and a copy of the survey instrument are presented in Appendix A.

newspaper or magazine articles; or workbooks. For the purposes of this analysis, the measure assumes that there was an average of 30 days in the prior month and that a student's report of having read each type of text represents a separate reading occurrence. Thus, the measure was constructed to allow for up to 210 self-reported school-related occurrences of reading activities during the prior month (7 survey items; Cronbach's alpha = .83).¹⁷

Amount of Non-School-Related Reading

This measure was constructed to reflect the self-reported number of times during the prior month that a student read each of seven different types of text outside school: fictional books; plays; poetry; (auto)biographies; books about science, technology, or history; newspaper or magazine articles; or reference books. For the purposes of this analysis, the measure assumes that there was an average of 30 days in the prior month and that a student's report of having read each type of text represents a separate reading occurrence. Thus, the measure was constructed to allow for up to 210 self-reported occurrences of reading activities outside school during the prior month (7 survey items; Cronbach's alpha = .73).

The third measure is intended to provide an indication of whether students use some of the skills and techniques that the ERO programs try to teach (asking questions of the text and reviewing and rereading). These strategies are second nature for proficient readers, and the measure can serve as a useful indicator of whether students are starting to incorporate them more explicitly into their reading behavior.

Use of Reflective Reading Strategies

This measure captures students' reported use of reflective reading strategies (each item is rated on a scale from 1 to 4) as they read for their English/language arts class and for one other academic class.¹⁸ Students were asked to rate their use of these two strategies on a scale from 1 (strongly disagree) to 4 (strongly agree) (4 survey items; Cronbach's alpha = .88).

The impact estimates for each of these three measures of reading behavior (amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies) are presented both in their original metrics and in effect-size units. Effect sizes provide an indication of the magnitude of the impact estimates relative to the variation in the measures for students in the study sample who were not exposed to the ERO programs. As with the

¹⁷Cronbach's coefficient alpha is a statistical measure of the degree to which the individual items used to create the multi-item construct are correlated with each other (Cronbach, 1951).

¹⁸The follow-up survey asked students to report on reading strategies that they use in social studies, science, and mathematics classes, if they are taking these courses. The measure relied on the social studies class, if the student reported taking social studies. Otherwise, it includes science. If the student was not taking either social studies or science, the measure includes mathematics.

test score outcomes, effect sizes are calculated as a proportion of the standard deviation of the given outcome for students in the non-ERO group.¹⁹ The standard deviation for the non-ERO group reflects the expected variability in the reading behavior that one would find in the absence of the ERO programs. The impact effect size, therefore, provides an indication of how much the ERO programs moved students along this variability in expected reading behavior.

Teacher Survey

The study team administered a two-part survey to ERO teachers during the summer training institutes held by the interventions' developers. Part 1 of the survey asked teachers about their backgrounds, their experiences with professional development activities, their school environments, and their beliefs about literacy instruction. Part 2 of the survey asked teachers about their impressions of the training they attended.

Implementation Data

Classroom Observations

The analysis of ERO program implementation fidelity in the first year of the study is based on field research visits to each of the 34 high schools during the second semester of the 2005-2006 school year. The primary data collection instrument for the site visits was a set of protocols for classroom observations and interviews with the ERO teachers.²⁰ The observation protocols provided a structured process for trained classroom observers to rate characteristics of the ERO classroom learning environments and the ERO teachers' instructional strategies. Each of these characteristics were selected for assessment because they were aligned with program elements specified by the developers and, by design, were aligned with supplemental literacy program elements that are believed to characterize high-quality interventions for struggling adolescent readers.²¹ Chapter 3 provides a more detailed description of the data collection process and a description of the summary measures of implementation fidelity that were developed from the classroom observation data. Appendix D provides further background on the properties of the classroom observation data and the fidelity measures.

¹⁹The standard deviation of the "amount of school-related reading" for the non-ERO group is 43.867. The standard deviation of the "amount of non-school-related reading" for the non-ERO group is 31.834. The standard deviation of the "use of reflective reading strategies" for the non-ERO group is 0.670.

²⁰The observation protocols can be found in Appendix D.

²¹Biancarosa and Snow (2004).

Teacher Interviews

During the field visits, the study team interviewed the ERO teacher using a semi-structured interview protocol that focused on teachers' perceptions of aspects of the intervention, of the coaching and support that they received from the developers, of the ease of implementing the program, and of students' responses to and challenges with the program. The study team also interviewed English/language arts teachers and elective teachers in order to explore the extent to which literacy instruction may be taking place in classes other than ERO.

Interviews with District Coordinators

The study team interviewed the ERO district coordinators during the site visits, to gather information as to their perceptions about implementing the program.

ERO Class Attendance Records

Each of the ERO teachers provided monthly school attendance data for all students in the study sample and ERO class attendance data for those students assigned to an ERO class.

Student Course Schedules

Each school provided the study team with copies of the schedules for all students in the study sample. One purpose of the schedule data is to confirm that ERO students were enrolled in the ERO classes and that non-ERO students were not.²² These data allow the study team to check for possible contamination — that is, for non-ERO students receiving the ERO program.

Follow-Up Data Collection and Response Rates

The follow-up GRADE assessment and survey were administered to students in the study sample late in the 2005-2006 school year. Overall, the follow-up data are available for 83 percent of the study sample. Table 2.3 shows that the response rate for students in the ERO group is 84 percent, compared with 81 percent for the non-ERO group. This difference is statistically significant (p-value less than or equal to 5 percent). Although the response rates for students in the ERO groups are similar for both the Reading Apprenticeship and the Xtreme Reading schools, the rate is somewhat lower for students in the non-ERO group from the Reading Apprenticeship schools. The difference in response rates between the ERO and non-ERO

²²See Chapter 4 for discussion of student schedules and enrollment in the ERO classes.

The Enhanced Reading Opportunities Study

Table 2.3

**Response Rates of Students in Cohort 1
Full Study Sample**

	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
<u>All schools</u>				
Response rate (%)	84.1	81.1	2.9 *	0.037
Sample size	1,675	1,241		
<u>Reading Apprenticeship schools</u>				
Response rate (%)	84.6	79.3	5.2 *	0.011
Sample size	811	574		
<u>Xtreme Reading schools</u>				
Response rate (%)	83.6	82.7	0.9	0.649
Sample size	864	667		

SOURCES: MDRC calculations from the Enhanced Reading Opportunities baseline data and follow-up GRADE assessment.

NOTES: This table represents the response rates for the follow-up GRADE assessment, which was administered in spring 2006 at the end of students' ninth-grade year. The follow-up student questionnaire was also administered at that time. The difference in response rates between the test and survey is negligible.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The p-value is the probability that the observed difference is the result of chance and does not represent a true difference between groups. The lower the p-value, the less confidence that there is not a difference between the two groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

groups in the Reading Apprenticeship schools is statistically significant (p-value less than or equal to 5 percent).²³

When response rates are less than 100 percent or when there are differences between program and control groups, it is important to investigate two concerns. First, does the respondent sample differ from the full study sample and from the nonrespondent sample? Second, within the respondent sample, are the ERO group and the non-ERO group still equivalent?

The ERO study team conducted a nonresponse analysis by examining differences in background characteristics between respondents and nonrespondents in the study

²³See Appendix Table B.1 in Appendix B.

sample.²⁴ While the respondent sample reflects the general characteristics of the full study sample, an overall F-test comparing the respondents and nonrespondents indicates that there are systematic differences between them in student characteristics. Most notably, response rates are lower for students with characteristics associated with doing poorly in school. For example, response rates are lower among students who were overage for grade than for those students who were not likely to have been held back in a previous grade. There are also differences in response rates across the participating high schools. Overall, however, response rates are similar for the schools using the Reading Apprenticeship program (82 percent) and those using Xtreme Reading (83 percent). The overall differences between respondents and nonrespondents suggest that one should be cautious when generalizing findings from the first cohort follow-up respondent sample.²⁵

As noted earlier, the three percentage point difference in the response rates between the ERO group (84 percent) and the non-ERO group (81 percent) is statistically significant (p-value less than or equal to 5 percent). This raises a concern about whether respondents in the ERO group differ systematically from respondents in the non-ERO group. Table 2.4 shows the background characteristics of all 2,413 students in the first cohort follow-up respondent sample and provides a comparison between the ERO and non-ERO groups. Like Table 2.2 for the overall study sample, Table 2.4 shows a high degree of similarity between the respondents in the ERO and non-ERO groups across the baseline characteristics. A general F-test indicates that, overall, there are no systematic differences between the ERO and non-ERO group respondents.²⁶ This suggests that one may have a high degree of confidence that differences in outcomes between the two groups reflect impacts of the ERO programs rather than preexisting differences in background characteristics.

The characteristics displayed in Table 2.4 indicate that the typical follow-up respondent sample member was reading well below grade level at the start of ninth grade and that many students have characteristics associated with a risk of doing poorly in school. On average, students had the same reading comprehension composite score of about 86 standard score points, corresponding to the 5.2 grade level and to the 17th percentile nationally. Also, over 70 percent of the students in the follow-up respondent sample are Hispanic or black, and over 45 percent

²⁴See Appendix B for the results of the statistical analyses that were conducted to assess differences between respondents and nonrespondents. Results are presented for all the participating high schools together and, separately, for the groups of schools using Reading Apprenticeship and Xtreme Reading, respectively.

²⁵See Appendix F for results from supplemental impact analyses that include sampling weights to account for differences between respondents and nonrespondents. These results indicate very little difference between the weighted and unweighted impact estimates.

²⁶See Appendix B for the results of the statistical analyses that were conducted to assess differences between the ERO and non-ERO groups in the respondent sample.

The Enhanced Reading Opportunities Study

Table 2.4

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	32.7	33.0	-0.3	0.803
Black, non-Hispanic	42.9	43.6	-0.7	0.632
White, non-Hispanic	18.3	17.2	1.1	0.437
Other	6.2	6.2	0.0	0.999
Gender (%)				
Male	50.1	51.3	-1.3	0.542
Female	49.9	48.7	1.3	0.542
Average age (years)	14.8	14.7	0.0	0.103
Overage for grade ^a (%)	28.1	25.1	2.9	0.092
Language other than English spoken at home (%)	47.1	45.9	1.2	0.512
Language spoken at home missing (%)	6.7	7.2	-0.5	0.618
Mother's education level (%)				
Did not finish high school	17.0	16.8	0.2	0.891
High school diploma or GED certificate	25.5	24.7	0.8	0.641
Completed some postsecondary education	29.0	31.3	-2.2	0.229
Don't know	21.0	19.7	1.3	0.426
Missing	7.5	7.7	-0.1	0.885
Father's education level (%)				
Did not finish high school	16.7	16.9	-0.2	0.894
High school diploma or GED certificate	22.9	22.1	0.8	0.645
Completed some postsecondary education	18.2	22.1	-3.9 *	0.015
Don't know	33.9	30.0	4.0 *	0.038
Missing	8.2	8.9	-0.7	0.518
GRADE reading comprehension ^b				
Average standard score	85.9	86.2	-0.3	0.143
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	34.4	37.0	-2.6	0.193
5.0 - 5.9 grade equivalent (%)	29.3	26.4	2.9	0.115
4.0 - 4.9 grade equivalent (%)	36.2	36.6	-0.3	0.859
Sample size	1,408	1,005		

(continued)

Table 2.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

reported that a language other than English is spoken in their homes. Tables 2.5 and 2.6 present similar results for students in the follow-up respondent samples from the Reading Apprenticeship schools and from the Xtreme Reading schools, respectively.

The similarity between the student characteristics of the follow-up respondent sample and full study sample — as well as the lack of systematic differences between the ERO and non-ERO groups in the follow-up respondent sample — indicate that the follow-up respondent sample preserves the balance that was achieved with random assignment for the full study sample. This balance was also preserved in the groups of schools using each of the two supplemental literacy programs.

Analytic Methods and Procedures

When examining the effectiveness of the ERO programs in improving students' reading achievement and behaviors, it is important to distinguish between measures of program “outcomes” and measures of program “impacts.” *Outcomes* refer to the measures of student performance, behaviors, achievement, and attitudes — in this case, reading achievement and reading behaviors at the end of the ninth-grade year. An *impact* is the effect that the ERO programs have on an outcome. The average outcome levels for students in the ERO group alone provide potentially misleading conclusions. Reading achievement and behaviors are likely to change for students for reasons not related to a special intervention like the ERO programs. In order to determine the net effect, or “value added,” of the ERO programs, it is necessary to compare the experiences of a group of students who were exposed to the ERO classes with a similar group of students who also applied but were not selected to enroll. As discussed earlier in this chapter, the ERO and non-ERO groups participating in this study were determined through a random

The Enhanced Reading Opportunities Study

Table 2.5

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample,
Reading Apprenticeship Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	31.5	31.8	-0.3	0.885
Black, non-Hispanic	43.1	43.8	-0.6	0.787
White, non-Hispanic	18.5	18.7	-0.2	0.916
Other	6.9	5.8	1.1	0.447
Gender (%)				
Male	50.0	51.7	-1.7	0.569
Female	50.0	48.3	1.7	0.569
Average age (years)	14.7	14.7	0.0	0.253
Overage for grade ^a (%)	27.0	25.2	1.8	0.475
Language other than English spoken at home (%)	45.0	45.1	0.0	0.991
Language spoken at home missing (%)	7.1	7.7	-0.5	0.701
Mother's education level (%)				
Did not finish high school	17.9	16.4	1.6	0.488
High school diploma or GED certificate	25.2	23.6	1.6	0.527
Completed some postsecondary education	27.8	30.1	-2.3	0.395
Don't know	21.3	21.7	-0.4	0.855
Missing	7.7	8.2	-0.5	0.740
Father's education level (%)				
Did not finish high school	16.8	17.1	-0.3	0.880
High school diploma or GED certificate	21.6	23.3	-1.7	0.491
Completed some postsecondary education	17.5	19.5	-2.0	0.395
Don't know	35.7	30.1	5.6 *	0.050
Missing	8.5	10.0	-1.6	0.331
GRADE reading comprehension ^b				
Average standard score	86.0	86.1	0.0	0.878
<i>Corresponding grade equivalent</i>	5.2	5.2		
<i>Corresponding percentile</i>	17	17		
6.0 - 7.0 grade equivalent (%)	36.4	35.5	0.9	0.743
5.0 - 5.9 grade equivalent (%)	29.0	28.0	1.0	0.712
4.0 - 4.9 grade equivalent (%)	34.5	36.5	-2.0	0.495
Sample size	686	454		

(continued)

Table 2.5 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

assignment process. The non-ERO group serves as a benchmark or counterfactual for how students in the ERO group would have performed if they had not had access to the programs. Therefore, the impacts (differences in outcomes between the ERO and the non-ERO groups) represent the effect that the ERO programs had students' reading achievement and other outcomes over and above what the students would have achieved had they stayed in their regularly scheduled elective class.

This section of the chapter discusses several technical issues that lie at the heart of the evaluation's capacity to produce valid and reliable estimates of the literacy interventions' impacts on student reading achievement and other outcomes. It first reviews the study's sample sizes and the implications for statistical power (that is, the precision with which the analysis can measure program impacts). The section then reviews the estimation model being used to generate impacts and finally discusses the standards used for indicating statistical significance (that is, the confidence one may have that the impact estimates are not zero).

Sample Sizes and Statistical Power

To ensure that the ERO impact evaluation could produce valid and reliable findings, several design features were put in place to enable the study to measure program effects (if they exist) that are large enough to be both meaningful in students' lives and relevant to policy debates about the efficacy of supplemental literacy interventions.²⁷ The number of schools and the number of student sample members are crucial factors that determine the degree to which the impacts on student achievement and other outcomes can be estimated with enough precision to

²⁷Appendix C provides a more detailed assessment of the statistical power of the ERO study's impact design and discusses the role of other design features and assumptions, including the use of pre-random assignment characteristics to improve precision and assumptions about fixed versus random effects.

The Enhanced Reading Opportunities Study

Table 2.6

Characteristics of Students in Cohort 1
Follow-Up Respondent Sample,
Xtreme Reading Schools

Characteristic	ERO Group	Non-ERO Group	Difference	P-Value for the Difference
Race/ethnicity (%)				
Hispanic	33.8	34.2	-0.4	0.838
Black, non-Hispanic	42.7	43.5	-0.8	0.686
White, non-Hispanic	18.0	15.9	2.2	0.239
Other	5.5	6.5	-1.0	0.463
Gender (%)				
Male	50.1	51.0	-0.9	0.762
Female	49.9	49.0	0.9	0.762
Average age (years)	14.8	14.7	0.0	0.244
Overage for grade ^a (%)	29.1	25.2	3.9	0.104
Language other than English spoken at home (%)	49.0	46.8	2.3	0.365
Language spoken at home missing (%)	6.4	6.7	-0.4	0.749
Mother's education level (%)				
Did not finish high school	16.1	17.1	-1.0	0.627
High school diploma or GED certificate	25.8	25.6	0.1	0.959
Completed some postsecondary education	30.2	32.3	-2.1	0.395
Don't know	20.6	17.8	2.8	0.197
Missing	7.3	7.2	0.2	0.893
Father's education level (%)				
Did not finish high school	16.6	16.7	-0.1	0.968
High school diploma or GED certificate	24.2	21.2	3.0	0.206
Completed some postsecondary education	18.8	24.4	-5.6 *	0.013
Don't know	32.3	29.7	2.6	0.322
Missing	8.0	7.9	0.1	0.930
GRADE reading comprehension ^b				
Average standard score	85.7	86.3	-0.5	0.058
<i>Corresponding grade equivalent</i>	5.1	5.2		
<i>Corresponding percentile</i>	16	17		
6.0 - 7.0 grade equivalent (%)	32.5	38.2	-5.6 *	0.036
5.0 - 5.9 grade equivalent (%)	29.6	25.1	4.5	0.068
4.0 - 4.9 grade equivalent (%)	37.8	36.8	1.1	0.690
Sample size	722	551		

(continued)

Table 2.6 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities baseline data.

NOTES: Baseline data were collected in fall 2005 at the start of the ninth-grade year.

The differences are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is the ERO group value minus the difference.

A two-tailed t-test was used to test differences between the ERO and non-ERO groups. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

^bThe national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.

An important goal for the design of the ERO study was to ensure that the sample sizes would be sufficient to allow for estimates of even “small” impacts on reading test scores and other outcomes both overall and for each of the supplemental literacy programs separately.²⁸ As discussed above, there are a total of 2,413 students in the Cohort 1 follow-up respondent sample for the impact analysis presented in this report. This includes 1,140 students from the 17 high schools using the Reading Apprenticeship program and 1,273 students from the 17 high schools using the Xtreme Reading program.

The overall study sample is equipped to detect impacts as small as 0.06 standard deviation units (referred to as “effect sizes”).²⁹ These pooled impact estimates provide insight into the impact of the family of interventions that share characteristics with Reading Apprenticeship and Xtreme Reading. The samples for each of the two supplemental reading programs are equipped to detect impacts as small as approximately 0.10 effect size.

²⁸There are no universally agreed-upon standards for what constitutes “small” versus “large” impacts. Some attempts have been made to examine the range of effects that have been found across a wide array of evaluations and to divide this range into segments that reflect the higher, middle, and lower categories of effects (see Lipsey, 1990). More recent work has begun to examine actual year-to-year rates of growth on a variety of achievement measures for students in a range of school districts and with a variety of background characteristics (see Bloom, Hill, Black, and Lipsey, 2006). These analyses provide additional background for interpreting the impact of interventions like those in the ERO study within the context of the expected growth in student outcomes nationally and under similar conditions.

²⁹The actual precision of estimated impacts may differ somewhat from those calculated for the statistical power analyses presented in Appendix C. These differences are due to such factors as actual variation in sample sizes, random assignment ratios, pretest scores, and outcomes levels across sites.

Statistical Model for Estimating Impacts

The ERO study impact analysis uses the following statistical model to estimate impacts on both reading achievement and reading behaviors:

$$Y_i = \sum_n \gamma_{0n} S_{ni} + \gamma_1 Y_{-i} + \sum_S \gamma_{2s} X_{si} + \beta_0 T_i + \varepsilon \quad (1)$$

Where:

Y_i = reading achievement or reading behaviors outcome for student i

$\sum_n S_{ni}$ = school dummy variable, one if student i is in school n and zero otherwise

Y_{-i} = the GRADE reading comprehension test score for student i before random assignment

$\sum_s X_{si}$ = other pre-random assignment characteristics for student i

T_i = one if student i is assigned to the ERO group and zero otherwise

ε_i = student-level random error term

In this model, β_0 represents the estimated impact of the ERO programs on the outcome of interest (Y_i). β_0 is a fixed-effect impact estimate that addresses the question: What is the impact of the ERO programs for the average student in the follow-up respondent sample? This approach is taken because this study most closely reflects an efficacy study of the effects of a new supplemental literacy intervention under relatively controlled conditions. Also, the sites and students were not selected to be a random sample of a larger population of sites. Instead, sites were selected purposively through the OVAE special SLC grant competition using specific criteria that differentiated these schools and districts from others that were not awarded a grant. In short, the impact estimates are not statistically generalizable to a larger population of districts, high schools, or students. As discussed above, however, on average, the participating schools share characteristics of other low-performing urban high schools across the country.

Equation 1 includes indicator variables for each of the participating high schools. These covariates capture a central feature of the study design in which random assignment was conducted within each of the participating high schools. These covariates are included to account for variation in the mean value of the dependent variable across the participating high schools.

Equation 1 also includes a covariate for each student’s GRADE reading comprehension test score at baseline and a covariate indicating whether the student is overage for grade (and likely to have been retained in a prior grade). These covariates are included to improve the precision of the impact estimates.

Statistical Significance

Equation 1 is estimated using ordinary least squares (OLS) regression, and a two-tailed t-test is used to assess the statistical significance of the impact estimate (β_0). *Statistical significance* is a measure of the degree of certainty one may have that some non-zero impact actually occurred. If an impact estimate is statistically significant, then one may conclude with some confidence that the program really had an effect on the outcome being assessed. If an impact estimate is not statistically significant, then the non-zero estimate is more likely to be a product of chance. For the purposes of this report, statistical significance is indicated in the tables by an asterisk (*) when the p-value of the impact estimate is less than or equal to 5 percent.

When making judgments about statistical significance, it is also important to recognize potential problems associated with conducting multiple hypothesis tests. Specifically, the analysis should avoid concluding that an impact estimate is statistically significant when in fact, there is no true impact (that is, relying on false positive results.) Likewise the analysis should not be so conservative with respect to producing false positives that it unduly increases the likelihood of missing true impacts when they exist (that is, relying on false negative results).

The statistical significance of the impact estimates presented in this report should be interpreted in light of two sets of safeguards aimed at attenuating the risk of drawing inappropriate conclusions about program effectiveness on the basis of ancillary hypothesis tests or statistically significant results that may occur by chance.³⁰ The first safeguard was to confine the analysis to a parsimonious list of outcome measures and subgroups. The shorter this list, the fewer the number of hypothesis tests and, thus, the less exposed the analysis will be to “spurious statistical significance” as a result of having tested multiple hypotheses. The primary evidence of overall ERO program effectiveness for this report will be reflected by estimates of program impacts on reading comprehension test scores (expressed in standard score values) for the full study sample and for each of the two ERO programs being evaluated. Vocabulary knowledge and student reading behaviors, while targets of the interventions and important to students’ literacy development, are considered secondary indicators of program effectiveness. Similarly, subgroups of students and subgroups of schools provide useful information about the relative impact of supplemental literacy programs, but they too are considered secondary indicators of effectiveness in this report.

³⁰See Appendix E for a more detailed discussion of the approach used to address the risks associated with multiple hypothesis tests.

The second safeguard uses composite statistical tests to “qualify” or call into question multiple hypothesis tests that are statistically significant individually but that may be due to chance in the context of mixed results.³¹ In general, these qualifying statistical tests estimate impacts on composite indices that encompass all the measures in a given domain or estimate the overall variation in impacts across subgroups.³² If the results of these tests are not statistically significant, this indicates that the statistical significance of the associated individual impact estimates *may have* occurred by chance. In these cases, the discussion of the impacts includes cautions or qualifiers about the robustness of the individual findings.

Finally, statistical significance does not directly indicate the magnitude or importance of an impact estimate — only the probability that an impact may have occurred by chance. Some statistically significant impacts may not be seen as policy relevant or as justifying the additional costs and effort to operate the programs under study. As a result, it is sometimes useful to frame the impact estimates in terms of other benchmarks and contexts, such as improvements found for related constructs or interventions, cost-effectiveness indicators, achievement gaps, or performance standards, which can help policy makers, practitioners and researchers gauge the importance or relevance of the findings. By the same token, lack of statistical significance for an impact estimate does not mean that the impact being estimated equals zero. It only means that the estimate cannot be distinguished from zero reliably. This can be due to the small magnitude of the impact estimate, the limited statistical power of the study, or some combination of both.

³¹Measurement of overall effects has its roots in the literature on meta-analysis (see O’Brien, 1984; Logan and Tamhane, 2003; and Hedges and Olkin, 1985). For a discussion of qualifying statistical tests to account for the risk of Type I error, see Duflo, Glennerster, and Kremer (2007). Other applications of these approaches are discussed in Kling and Liebman (2004) and Kling, Liebman, and Katz (2007).

³²See Appendix E for a more detailed description of the method used to conduct these qualifying statistical tests. Appendix E also includes tables with the results of these tests.

Chapter 3

Implementing the Supplemental Literacy Programs

This chapter describes the two supplemental literacy programs that are being used in the high schools participating in the Enhanced Reading Opportunities (ERO) study and assesses the fidelity of their implementation during the first year of the study. The chapter's first section provides an overview of the process used to select the programs at the start of the study and then describes the programs' core elements as presented in the proposals submitted by their developers and in other literature and materials associated with the programs. The second major section of the chapter presents the background characteristics of the teachers who elected to teach the ERO classes and describes the training activities and technical support they received to prepare them for this work. The third section of the chapter discusses findings on the fidelity with which each of the supplemental literacy programs was implemented in the participating high schools. The chapter concludes with a discussion of factors affecting the first year of implementation and how the second year of implementation has been different.

There are several key points made in this chapter:

- The two programs evaluated were selected by an independent national panel of adolescent literacy experts from among 17 proposals through a competitive process.
- Both programs focus on establishing a positive learning environment in the classroom to facilitate the delivery of instruction in reading comprehension processes and strategies. The comprehension instruction seeks to make explicit the processes used by capable readers, teaching less proficient students to pay attention to *how* they read so that they can improve their comprehension of *what* they read.
- Teachers self-selected to teach the ERO programs and were approved by the schools, districts, and ED. They held a high school teaching license or certificate and had an average of over 11 years of teaching experience. Three of the 34 starting teachers discontinued their involvement in the study before the end of the school year, and their replacements were trained and provided with coaching as they took over the ERO classes.
- The implementation of the ERO programs in 16 of the 34 participating high schools was deemed to be “well aligned” with the respective program models in the first year. Eight of the schools were found to have achieved a level of

implementation “moderately aligned” with both the classroom learning environments and the reading comprehension instruction practices specified by the developers. Implementation of the ERO programs in the remaining 10 high schools was found to be problematic, and either the classroom learning environments and/or the comprehension instruction practices were deemed to be “poorly aligned” with the models specified by the developers.

Characteristics of the Supplemental Literacy Programs: Reading Apprenticeship Academic Literacy and Xtreme Reading

The supplemental literacy programs were selected through a competitive proposal process that was managed by the study team and guided by a panel of seven nationally known experts in adolescent literacy research and program development. A request for proposals (RFP) was advertised in a wide range of education publications and was disseminated to over 40 organizations that develop and implement high school curricula.¹ The RFP specified that prospective supplemental literacy programs must be research-based, high-quality programs that provide instruction in the areas that experts increasingly agree are necessary for effective adolescent literacy instruction, as outlined in *Reading Next*, but that were not yet rigorously tested.² The prospective programs were to have been developed already (that is, not be new programs) and to be ready for systematic use in multiple schools and districts.

Seventeen proposals were submitted in response to the RFP. After a review of the research base presented in the proposals for each program, the proposals were rated by the panel of adolescent literacy experts. The developers of four of the proposed programs were invited to give oral presentations before the panel, staff from the U.S. Department of Education (ED), and the ERO study team. Based on the presentations and subsequent discussion, the panelists recommended and ED accepted two programs for inclusion in the study: WestEd’s Reading Apprenticeship Academic Literacy and the University of Kansas Center for Research on Learning’s (KU-CRL) Xtreme Reading.

Overall Goals and Approach

The overarching goal of both Reading Apprenticeship and Xtreme Reading is to help students adopt the strategies and routines used by proficient readers, improve their comprehension skills, and motivate them to read more and enjoy reading. Both programs emphasize the importance of establishing a specific type of learning environment in the classroom that is con-

¹American Institutes for Research (2004).

²Biancarosa and Snow (2004).

ducive to the effective delivery of the core instructional strategies by the teacher and to facilitate student and teacher interactions around the reading skills that are being taught and practiced. They both use a “cognitive apprenticeship” approach to instruction in which the teacher initially takes the lead in modeling the strategies that proficient readers use and then gradually increases the responsibility of the students to demonstrate and apply these strategies. The teachers seek to make explicit *how* proficient readers read, and they support their students in recognizing and using the strategies or methods used by stronger readers. That is, both programs focus students’ attention on *how* they read (a metacognitive process) to help the students better understand *what* they read (understanding content). Also, both programs integrate direct, whole-group instruction with small-group and individualized instruction.³

Key Components

The key components of Reading Apprenticeship and Xtreme Reading are discussed categorically below. This discussion is based on information provided by the two program developers. Table 3.1 also presents these components by category. These components are the specific aspects of the programs’ instructional approaches that the developers expect to improve the literacy skills of high school students.⁴

Developer’s Implementation Philosophy

In implementing Reading Apprenticeship, teachers are guided by the concept of “flexible fidelity.” That is, while the program includes a detailed curriculum, the teachers are trained to adapt their lessons to meet the needs of their students and to supplement program materials with readings they expect to be motivating to their classes. Teachers have flexibility in how they include various aspects of the Reading Apprenticeship curriculum in their day-to-day teaching activities, but have been trained to do so such that they maintain the overarching spirit, themes, and goals of the program in their instruction.

Xtreme Reading was developed with the philosophy that the presentation of instructional material — particularly the order and manner in which the material is presented — is of critical import to the students’ understanding of it, and as such teachers are trained to deliver course content and materials in a precise, organized, and systematic fashion designed by the

³Additional information about the Reading Apprenticeship Academic Literacy course is available on the Internet at <http://www.wested.org/cs/we/view/serv/111>; information about the Xtreme Reading course is available at <http://www.xtremereading.org/>. Furthermore, the descriptive material about the program-specific observation rating scales in Appendix D provides more information specific to each program.

⁴The proposals submitted by the two developers, WestEd (2004) and University of Kansas (2004), contain information about the key components of their programs. These proposals are unpublished and cannot be released based on the rules of the competition through which the programs were selected.

The Enhanced Reading Opportunities Study

Table 3.1

Key Components of the ERO Programs

	WestEd/Reading Apprenticeship	KU-CRL/Xtreme Reading
Developer’s Implementation Philosophy	“Flexible fidelity” guided by the instructional and behavioral/social needs of the students	Prescribed daily lesson plans and time limits on classroom activities
Role of Teacher	Instructor as “master reader,” apprenticing students in various literacy competency areas and drawing on variety of materials	Instructor explicitly teaches seven reading strategies using a prescriptive eight-stage instructional approach with step-by-step instructional materials
Curriculum Design	<p><u>Learning Environment</u> Establish “social reading community” early in program</p> <p><u>Comprehension Instruction</u> Five curricular strands of classroom instruction:</p> <ol style="list-style-type: none"> 1. Metacognitive Conversation 2. Silent Sustained Reading 3. Language Study 4. Content/Theme 5. Writing 	<p><u>Learning Environment</u> Focus at beginning of course on teaching social and behavioral skills and strategies aimed to develop a productive and positive classroom learning environment</p> <p><u>Comprehension Instruction</u> Focus of rest of course on developing literacy skills through seven learning strategies:</p> <ol style="list-style-type: none"> 1. LINC’S Vocabulary Routine 2. Word Mapping 3. Word Identification 4. Self-Questioning 5. Visual Imagery 6. Paraphrasing 7. Inferencing
Teaching Strategies	Instructors usually use one or two of the following routines during class period: <ol style="list-style-type: none"> 1. Think aloud 2. Talking to the text 3. Metacognitive logs/journals 4. Preambles (daily warm-ups) 	Each strategy is taught using a prescribed eight-stage instructional methodology: <ol style="list-style-type: none"> 1. Describe 2. Model 3. Verbal practice 4. Guided practice 5. Paired practice 6. Independent practice 7. Differentiated instruction 8. Integration and generalization
Program Type	Supplemental course, like an elective	Supplemental course, like an elective
Duration	One school year	One school year

developers. Xtreme Reading teachers follow a prescribed implementation plan, following specific day-by-day lesson plans in which activities have allotted segments of time within each class period. However, there are opportunities in the Xtreme Reading instructional program for teachers to use responsive instructional practices to adapt and adjust to student needs that arise as they move through the highly structured curriculum.

Role of Teacher

Both Reading Apprenticeship and Xtreme Reading are grounded in the principle of a cognitive apprenticeship. That is, the teacher assumes the role of reading expert whose task is to share expertise in explicit ways with the students and then to support their development of those skills and nurture their increased independence in using them. The process is one that starts off as teacher-centered and gradually transitions to being student-centered. In Reading Apprenticeship — where the teacher is considered the “master reader” for the students, who are the “reading apprentices” — the transition is facilitated through the teacher’s integration of the four dimensions of classroom life (personal, social, cognitive, and knowledge-building; described below), which he or she links together through ongoing metacognitive conversations (thinking internally and talking externally about reading processes).

For the Xtreme Reading teacher, this transitional process is a specific eight-stage instructional model through which seven specific literacy strategies are taught. In Xtreme Reading classes, the expectation is that the learning of each strategy begins with specific teacher-directed instruction and that control is relinquished to students incrementally as they progress through the stages. By the eighth stage, students are working independently and have an understanding of the application of the strategy outside the Xtreme Reading classroom.

Curriculum Design and Teaching Strategies

As discussed above, the two programs are attentive to both the learning environment in the classroom and the nature of the literacy instruction, particularly around reading comprehension. The curriculum design and the teaching strategies of the two ERO programs reflect these two priorities. Table 3.1 provides an overview of the key elements of each ERO program. The developers’ curriculum designs both highlight the equal importance of creating a conducive classroom learning environment and focusing instruction on strategies that promote reading comprehension skills and proficiency.

The core of the Reading Apprenticeship program is the integration of four dimensions: social, personal, cognitive, and knowledge-building. The social and personal dimensions reflect the attention of the program to the *learning environment* for the class. The social dimension refers to adolescents’ interests in peer interaction and in larger social, political, and cultural issues. The personal component addresses students’ own goals for reading and for reading improve-

ment. These aspects of the program are combined in the establishment of a social reading community, a classroom environment that allows for the respectful, open exchange of ideas considered essential for the program to have effective comprehension instruction.

The cognitive and knowledge-building dimensions are the *instructional components* of the Reading Apprenticeship program. They address students' needs to increase both their repertoire of comprehension strategies and their background knowledge, expanding their knowledge base through reading, and providing knowledge about aspects of strong reading such as word construction, vocabulary, text structure, or figurative language. The instructional components are delivered across the following three major thematic units during the school year: "Who Am I as a Reader?" "Reading History," and "Reading Science and Technology." Within each unit, the teacher incorporates the five key curricular strands of the program:

- **Metacognitive conversations.** The students and the teacher think and talk about the thinking processes that are engaged when reading.
- **Silent sustained reading.** The student reads a book of his or her choice for 20 to 25 minutes at least twice a week to build reading fluency, comprehension, motivation, and stamina.
- **Language study.** The teacher and the students routinely practice strategies and learn skills at the word, sentence, and text levels to enhance language development.
- **Content/theme.** The teacher uses the majority of instructional time to address one of the three thematic units of the curriculum so that students are able to apply what they are learning in the classroom to their other classrooms and relate what they are learning to contexts other than Reading Apprenticeship.
- **Writing.** The teacher provides opportunities for the students to write and provides new knowledge of writing processes and strategies as needed.

The curriculum strands are taught and reinforced through the use of four teaching strategies: *think alouds*, *talking to the text*, *metacognitive logs*, and *daily preambles*. These strategies offer teachers and students opportunities to interact around what they are reading and how they are reading.

The Xtreme Reading program also emphasizes creating a positive *learning environment* in the classroom. The program aims to create a structured classroom climate with explicit social and behavioral expectations and regular routines for both students and teachers. The main tenet of classroom management is time-on-task behavior; this is essential to successful implementa-

tion of the instructional sequence. Student motivation and engagement are encouraged through several activities that help students set short- and long-term goals for their learning and through the availability and sharing of high-interest novels about students who have overcome academic obstacles. Teachers seek to help students to set real purposes for learning and to link their learning to personal goals.

The program's *literacy instruction* involves both a systematic component (driven by the curriculum) and a responsive component (driven by student needs). The systematic component involves teaching seven reading strategies following lesson plans provided by the developer that map out daily instruction. Two strategies focus explicitly on vocabulary: LINC and Word Mapping. Five strategies focus more directly on comprehension: Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inferencing. Each strategy is taught using an eight-stage model that starts off being highly teacher-centered (the teacher describes and models the strategy in the first two stages), to being shared work between the teacher and the students (verbal and guided practice), to being more and more the responsibility of the students (paired practice between students and independent student practice). The seventh stage is differentiated instruction, allowing those struggling with the strategy to receive additional support and those who have been successful learning the strategy more and varied opportunities for practice. The eighth stage, integration and generalization, involves students' taking the strategy beyond the Xtreme Reading classroom and materials and applying it to reading in other classes. The responsive instruction component focuses on assessing and addressing individual student needs as they arise. The responsive instruction component represents where flexibility enters into Xtreme Reading instruction.

Both ERO programs were developed from preexisting programs prior to implementation in the ERO study. The program developers adapted their already existing curricula to create programs that would be supplemental, yearlong reading classes. The Reading Apprenticeship Academic Literacy curriculum combined elements of two WestEd programs, Reading Apprenticeship and Academic Literacy. These programs had been the focus of most of the work within WestEd's Strategic Literacy Instruction initiative. Instruction in Reading Apprenticeship helps students identify weaknesses in their reading skills and improve them through mastering and then consciously applying advanced reading strategies. Academic Literacy is usually woven into content-area instruction so that students learn to apply subject-specific skills and strategies in areas such as science and social studies. The curriculum used in this study offered instruction in strategic reading within three themed units, two of which emphasized content-area reading. The Xtreme Reading curriculum combined the components of the Strategic Instruction Model (SIM) for reading improvement that has been developed, studied, and refined at the University of Kansas Center for Research on Learning for close to 30 years. SIM content consists of six specific reading processes, such as vocabulary identification and strategies for making inferences from the text. Previous implementation of SIM had followed the eight-stage instructional

model used in Xtreme Reading but had not combined the six reading strategies into a full-year curriculum for use in self-contained intervention classes. Further, two versions of this curriculum were developed to accommodate both 45- and 90-minute instructional blocks.

The ERO Teachers and Their Preparation for the ERO Programs

Teachers play a key role in both programs selected for the study. The study sought to have experienced, core-content-area teachers implement the programs and to provide adequate training and support for them. The teachers were nominated by their schools on the grant applications submitted to the Office of Vocational and Adult Education (OVAE) at ED. Additionally, participating districts and schools committed to make these teachers available for professional development activities prior to the start of the school year and on an ongoing basis during the year.

Teacher Characteristics

The Request for Proposals from OVAE to which school districts responded in their application for grant funding and participation in this study specified that teachers selected to teach the ERO classes at each high school should have at least two years of experience and be certified core-content-area teachers — specifically, English or social studies teachers — and not necessarily reading specialists. The project sought to target content-area teachers rather than reading teachers to teach the classes in order to enhance the replicability of the interventions if they proved to be effective. First, the study sought to demonstrate that if content-area teachers could be trained to deliver a literacy program, schools and districts that later chose to pursue this type of intervention may have a more realistic chance to identify staff to teach it without being restricted to reading specialists. Second, one of the goals of both interventions is transference — helping students use the literacy skills that they develop in their content-area classes. Thus, it was hoped that involving content-area teachers would help facilitate this.

Table 3.2 provides a list of background characteristics for the teachers in each of the two ERO programs.⁵ The average number of years of previous experience for ERO teachers was 11.2 years, although prior teaching experience ranged from student teaching to over 30 years as a regular classroom teacher. Almost three-quarters (73.5 percent) of the teachers had graduate-level degrees, and almost all (97.1 percent) held high school-level certification. The majority of the teachers (76.5 percent) were certified in English/language arts, with nearly 18

⁵Information in Table 3.2 is drawn from the survey that teachers completed at the beginning of the ERO training or at the beginning of their tenure as an ERO teacher. The information in the table reflects the characteristics of the teacher who spent the longest period of time as the ERO teacher in each participating school. Three of the teachers who began the 2005-2006 school year teaching the ERO students left that position before the end of the school year.

The Enhanced Reading Opportunities Study

Table 3.2

Background Characteristics of ERO Teachers

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Race/ethnicity (%)			
Black	20.6	23.5	17.7
White	67.7	64.7	70.6
Other	11.8	11.8	11.8
Gender (%)			
Male	23.5	11.8	35.3
Female	76.5	88.2	64.7
Total time teaching (years) ^a	11.2	9.0	13.5
Total time teaching at current school (years) ^b	4.8	4.7	4.9
Total time teaching at current level (years) ^a	7.1	5.7	8.6
Total time teaching English/language arts or social studies (years) ^a	10.4	8.4	12.7
Master's degree or higher (%)	73.5	70.6	76.5
Holds high school-level teaching certification (%)	97.1	100.0	94.1
Subject matter certification (%)			
Certified in English/language arts	76.5	70.6	82.4
Certified in social studies	17.7	23.5	11.8
Certified in other subject	5.9	5.9	5.9
Number of professional development workshops attended in the last two years ^a	3.8	4.2	3.3
Number of hours spent in professional development workshops during the last two years ^b	45.4	40.9	50.4
Taught the ERO class for the full school year (%)	91.2	100.0	82.4
Sample size	34	17	17

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study baseline teacher survey.

NOTES: For three schools, the original teacher was replaced during the school year. The table includes the teacher who spent the most time teaching the ERO program.

Rounding may cause slight discrepancies in calculating sums and differences.

^aMissing data: One to two teachers did not respond.

^bMissing data: Four to five teachers did not respond.

percent holding social studies certification and 6 percent holding certification in some other area. Teachers reported attending an average of 45.4 hours of professional development in the two years prior to the beginning of the ERO program.⁶

Training and Technical Assistance

Training and technical assistance were delivered to the ERO teachers in the following ways: Reading Apprenticeship teachers attended one 5-day summer training institute as well as two 2-day booster training sessions during the 2005-2006 school year. They also received ongoing support through three 2-day coaching visits during the year and access to a special online listserv that was set up for the project. Xtreme Reading teachers attended one 5-day summer training and one 2-day booster training during the year. They also received three 2-day on-site coaching visits. District coordinators were asked to attend the trainings to familiarize them with the programs in case they had to provide technical assistance or other support to ERO teachers. Table 3.3 summarizes the activities provided by each of the developers for the 2005-2006 school year.

The Enhanced Reading Opportunities Study

Table 3.3

Training and Technical Assistance Provided During the 2005-2006 School Year, by ERO Program

	Summer Training	School-Year Booster Training	Additional Supports
Reading Apprenticeship	One 5-day training (August)	Two 2-day trainings (November; February)	Three 2-day on-site coaching visits Weekly e-mail and phone calls Listserv
Xtreme Reading	One 5-day training (August)	One 2-day training (January)	Three 2-day on-site coaching visits Weekly e-mail and phone calls Additional technical assistance for replacement teachers

⁶Differences between teachers in each ERO program were not tested for statistical significance. There is one ERO teacher per school, which means that teacher characteristics are also school characteristics. As discussed later in the chapter, the impact analysis accounts for differences across school characteristics (and, thus, across teachers) by including regression covariates for each school.

Summer Trainings

The summer teacher training institutes for both programs were conducted in August 2005. The Reading Apprenticeship Academic Literacy training was conducted by the program developer, experienced Reading Apprenticeship teachers, and the two coaches who would work with the ERO teachers throughout the year. The Xtreme Reading training was conducted by the program developers, research staff from the University of Kansas Center for Research on Learning, and the coaches who would work with the teachers throughout the year. Each of the trainings provided the teachers with an introduction to the program as a whole but also included time focused on the curricular units to be taught during the first quarter of the course. Training methods across both summer institutes included modeling, discussion, and formal presentations as well as large-group and small-group activities. Teachers also had time to meet with the coaches with whom they would be working during the year. Fifteen of the 17 Reading Apprenticeship teachers attended the summer training. The other two attended national Reading Apprenticeship workshops before they began teaching the course.⁷ All of the Xtreme Reading teachers attended the summer training session.

Items on surveys administered to the teachers at the conclusion of the summer training probed the teachers' perceptions of their preparedness for teaching the ERO classes and their sense of the challenge they faced in implementing the programs. Thirty-three of the 34 ERO teachers (one teacher did not respond to the item) agreed or strongly agreed with the statement "I will be able to present this program confidently to students with the help of the manuals, other materials, and support of the professional developers." Additionally, 29 of the ERO teachers (15 of the 17 Reading Apprenticeship teachers and 14 of the 17 Xtreme Reading teachers) disagreed or strongly disagreed with the statement that the "[Reading Apprenticeship and Xtreme Reading]-recommended strategies and activities seem difficult to implement." Of the other five ERO teachers, two did not respond to the item and three agreed or strongly agreed that it would be difficult to implement the programs' strategies and activities.

Booster Trainings

The booster trainings during the school year (two for Reading Apprenticeship and one for Xtreme Reading) were conducted in a similar format to the summer training institutes and were two days each in duration. The program developers introduced the teachers to the curricular units coming up next in the programs as well as to the computer-based components of the

⁷The Reading Apprenticeship Academic Literacy course being implemented in the ERO Study is an adaptation of the preexisting Reading Apprenticeship program on which the national workshops were focused. While at the national workshops, these two ERO teachers received additional training that addressed aspects of Reading Apprenticeship that are specific to the ERO Study.

courses. Each of the trainings also provided time for the teachers to meet with their coaches and opportunities for the teachers and developers to discuss any issues with the implementation of the program that had come up during the first part of the year. All 17 of the Reading Apprenticeship teachers attended both booster training sessions. Sixteen of the Xtreme Reading teachers attended the booster training session in person, and one teacher participated by telephone.

Ongoing Technical Assistance

Both programs provided on-site coaching and electronic communication among teachers and their coaches. Reading Apprenticeship also made a listserv available to teachers. The Reading Apprenticeship and Xtreme Reading coaches made three 2-day visits to each of the teachers, during which they observed classes, modeled instruction, and in some cases co-taught lessons, in addition to working through issues that each teacher was experiencing. In the three instances of teacher turnover, coaches provided additional technical assistance to the replacement teachers.

Implementation Fidelity

This section of the chapter examines the fidelity with which the two supplemental literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — were implemented. In particular, it defines the method by which composite measures of implementation fidelity were computed for each school, based on classroom observations conducted by study team members during site visits in the second semester of the first year of implementation. In the context of this study, “fidelity” refers to the degree to which the observed operation of the ERO program in a given high school approximated the intended learning environments and instructional practices that were specified by the model developers.

Overall ratings of the implementation fidelity of the ERO programs at each school provide a context for interpreting the study’s impact findings and offer information to policymakers and practitioners about factors they may wish to consider if establishing these programs or ones like them in high schools.

Data Sources and Measures

As noted in Chapter 2, the analysis of ERO program implementation fidelity in the first year of the study is based on field research visits to each of the 34 high schools during the second semester of the 2005-2006 school year.⁸ The classroom observation protocols used in the site visits provided a structured process for observers to rate characteristics of the ERO class-

⁸Appendix D provides more detailed description of the site visits.

room learning environments and the ERO teachers' instructional strategies. The instrument included ratings for six characteristics (referred to as “constructs” from here forward) that are common to both programs and ratings for seven program-specific constructs. The analysis of the classroom observation ratings sought to capture the implementation fidelity of two key overarching dimensions of both programs: the classroom learning environment and the instructional strategies that focused on reading comprehension. A composite measure of implementation fidelity for each dimension was calculated from the average ratings for both general and program-specific constructs.

Table 3.4 provides a list of the constructs that were combined to create composite ratings for the learning environment and comprehension instruction dimensions, respectively, for the ERO programs in each high school. The *learning environment composite* was calculated as the average of ratings on two general constructs and ratings of one or two program-specific constructs for Reading Apprenticeship and Xtreme Reading, respectively. The *comprehension instruction composite* was calculated as the average of ratings on two general constructs and ratings of five program-specific constructs.⁹ The composite measures ranged from one to three and were rounded to the nearest tenth of a point.

Based on the composite ratings for each of the two program dimensions — learning environment and comprehension instruction — the implementation fidelity for each dimension was classified as “well aligned,” “moderately aligned,” or “poorly aligned” with the models specified by the program developers. The fidelity analysis focused on identifying schools where implementation of one or both of the two key program dimensions was especially problematic. This focus is particularly relevant to the first year of implementation, when the programs were new to the schools and the teachers and their lack of prior experience with the programs presented a more challenging implementation scenario. Thus, the definitions below for each level of implementation fidelity include not only information about average ratings but also the number of constructs rated in Category 1 — implementation that was poorly aligned with the expectations of the ERO programs.

Implementation fidelity for the learning environment or comprehension instruction dimensions was characterized as *well aligned* when the average rating across the relevant general and program-specific constructs was 2.0 or higher. That is, the school's ERO program was rated as moderately (a Category 2 rating) or well aligned (a Category 3 rating) with the program models on all or almost all of the constructs included in that dimension. As it turned out, the

⁹Note that, for Xtreme Reading, the program-specific component comprises two subcomponents: curriculum-driven or systematic instruction and needs-driven or responsive instruction. Appendix D provides a detailed description of the method used to average the ratings on individual constructs to create the composites for the two overarching program dimensions.

The Enhanced Reading Opportunities Study

Table 3.4

**Dimensions and Component Constructs of Implementation Fidelity,
by ERO Program**

Dimension	Component	Reading Apprenticeship	Xtreme Reading
Learning Environment	General Instructional Constructs	Classroom climate On-task participation	Classroom climate On-task participation
	Program-Specific Constructs	Social reading community	Classroom management Motivation and engagement
Comprehension Instruction	General Instructional Constructs	Comprehension Metacognition	Comprehension Metacognition
	Program-Specific Constructs	Metacognitive conversations Silent sustained reading Content/theme integration Writing Integration of curriculum strands	Curriculum-driven (systematic) instruction <ul style="list-style-type: none"> • Structured content • Research-based methodology • Connected, scaffolded, informed instruction Needs-driven (responsive) instruction <ul style="list-style-type: none"> • Student accommodations • Feedback to students

schools with fidelity rated as *well aligned* had no more than one construct for each implementation dimension rated in Category 1.

The key dimensions were designated as *moderately aligned* in terms of implementation fidelity if the average rating across the general and program-specific constructs used to create the relevant composite was within the range of 1.5 to 1.9. In these cases, the school's ERO program was observed to have some problems with implementation. In terms of learning environment, these schools had one construct rated in Category 1 (out of three or four constructs used to calculate the composite for Reading Apprenticeship or Xtreme Reading schools, respectively). On the comprehension instruction dimension, schools had three or fewer constructs rated in Category 1 (out of seven constructs used to calculate the composite score). These schools also met with some implementation success, with half or more of the constructs that make up the dimension being rated as moderately or well aligned with the program models.

The implementation fidelity of key program dimensions in a school was rated as *poorly aligned* when the average composite rating across the general and program-specific constructs fell below 1.5. This resulted when the school's ERO program was rated in Category 1 for half or more of the general or program-specific constructs that make up the dimension. These programs were the least representative of the activities and practices intended by the respective program developers.

The ratings and resulting categories indicate whether the programs reflected the characteristics of the classroom learning environments and instructional strategies intended by the developers. While it is reasonable to expect that higher fidelity programs could produce stronger impacts than programs where the fidelity was only a limited reflection of the intended model, other factors could intervene to make higher fidelity programs ineffective or to make limited or inadequate fidelity programs effective.

Findings

Table 3.5 provides a summary of the findings regarding implementation fidelity. The top two panels of the table provide a summary of the number of schools whose composite rating on the classroom learning environment and comprehension instruction dimensions fell into the well-aligned, moderately aligned, and poorly aligned categories of fidelity. The bottom panel of the table categorizes schools in terms of their overall implementation fidelity, based on their ratings across both implementation dimensions. The discussion that follows focuses first on each implementation dimension and then turns to overall fidelity, which accounts for the importance of the implementation of both dimensions to the ERO programs.

The Enhanced Reading Opportunities Study

Table 3.5

Number of ERO Classrooms Well, Moderately, or Poorly Aligned to Program Models on Each Implementation Dimension, by ERO Program

Implementation Dimension	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
<u>Learning environment</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	26	14	12
Moderately aligned implementation (composite rating is 1.5-1.9)	4	2	2
Poorly aligned implementation (composite rating is less than 1.5)	4	1	3
<u>Comprehension instruction</u>			
Well-aligned implementation (composite rating is 2.0 or higher)	16	7	9
Moderately aligned implementation (composite rating is 1.5-1.9)	9	4	5
Poorly aligned implementation (composite rating is less than 1.5)	9	6	3
<u>Combined dimensions</u>			
Well-aligned implementation on both dimensions	16	7	9
Moderately aligned implementation on at least one dimension and moderately or well-aligned implementation on the other dimension	8	4	4
Poorly aligned implementation on at least one dimension	10	6	4
Sample size	34	17	17

SOURCES: MDRC and AIR calculations from classroom observation data.

NOTES: Implementation with a composite score of less than 1.5 for a given dimension was deemed to be at the beginning stages of development. The implementation for these dimensions was designated as poorly aligned with the program models.

Implementation with composite scores between 1.5 and 1.9 for a given dimension exhibited at least moderate development in some areas while being at the beginning stages of development in other areas. The implementation for these dimensions was designated as moderately aligned.

Implementation with scores of 2.0 or higher for a given dimension exhibited well-developed fidelity on several areas and at least moderate development in most other areas. The implementation for these dimensions was designated as well aligned.

Fidelity by Implementation Dimension

As described earlier in the chapter, the first curriculum unit for both Reading Apprenticeship and Xtreme Reading focuses on the learning environment in the classroom. This focus involves setting expectations for the organization of the classroom, for how students should interact with the teacher and with their peers, and for the daily and weekly schedules of classroom activities. These same expectations are reinforced in each of the subsequent curriculum units. Table 3.5 shows that the ERO programs in 26 of the 34 high schools reached a level of implementation that was well aligned with the program models in terms of classroom learning environment dimension. Four schools were rated as demonstrating moderate alignment on this dimension, and four schools were rated as demonstrating poor alignment.

Compared with aspects of the ERO programs focused on the classroom learning environment, comprehension instruction evolves differently over the course of the year and varies across curriculum units. Although instructional strategies that focus on metacognition and content are incorporated in all the curriculum units, teachers were learning new instructional features of each ERO program continuously throughout the first year of implementation. As a result, it took a year of work with the ERO program for teachers to be exposed to and use the full repertoire of comprehension instruction strategies. As shown in Table 3.5, implementation was rated as well aligned on the comprehension instruction dimension for the ERO programs in 16 of the schools. Nine schools demonstrated moderate alignment, and nine schools demonstrated poor alignment, on the comprehension instruction dimension.

Differences in Fidelity, by Implementation Dimension

The pattern of findings shown in Table 3.5 indicates that more schools reached a level of well-aligned implementation fidelity on the learning environment dimension (26 schools) than on the comprehension instruction dimension (16 schools). Two hypotheses offer explanations for this observed difference in the fidelity achieved by schools on these two dimensions. First, this difference may reflect how these programs evolve during their implementation. The continuous and mutually reinforcing way that the elements of the classroom learning environment dimension are situated in the curriculum presents ongoing opportunities for teachers to refine their implementation of this dimension's elements and reach alignment with the program model. The elements of comprehension instruction are revealed in a more step-by-step way, unit by unit as the year progresses. Thus, teachers do not have the same continuous opportunity to refine their implementation of each instructional element. A second hypothesis for the difference in fidelity achieved on the two implementation dimensions is the difference in teachers' experience with teaching reading as opposed to developing a positive classroom environment. The instructional aspects of the programs were new to the teachers, who came to the program

predominantly from core-content-area teaching and not reading or literacy. However, the principles behind the learning environment dimension of the program models reflect principles often advocated for classrooms across subject areas, such as respect between individuals and creating a safe space for sharing opinions and ideas. The program developers emphasize the importance of both program dimensions, but it is useful for policymakers and practitioners to know that, in the implementation of these programs or similar ones, different aspects of the programs may develop more quickly than others.

Rating the Overall Fidelity of ERO Program Implementation

The bottom panel of Table 3.5 clusters schools based on their levels of implementation fidelity across both the classroom learning environment and the comprehension instruction dimensions. Because the classroom learning environments and comprehension instruction activities were designed to be interdependent and mutually reinforcing, the implementation of the ERO program in a given school was deemed to be well aligned with the program model overall only if both of these dimensions were rated in this category. The ERO programs in 16 of the 34 schools were found to have reached this level of implementation on both the classroom learning environment and the comprehension instruction dimensions. These schools did not necessarily represent exemplary versions of the ERO program model being used, although some of them did. While there is variation among these schools, the assessment of their implementation fidelity revealed that all constructs or all but one construct across both implementation dimensions were rated as either moderately (Category 2) or well aligned (Category 3) with the program models. These 16 schools include seven Reading Apprenticeship schools and nine Xtreme Reading schools.

In eight of the 34 high schools, the implementation of the ERO program was rated as moderately aligned with the program model for at least one of the two key program dimensions. It should be noted that, for these schools, neither of the dimensions was rated as poorly aligned. In fact, the classroom learning environment was rated as well aligned for the ERO programs in six of these schools, while the comprehension instruction was found to have reached a level of moderate alignment. In the remaining two schools, both the classroom learning environment and the comprehension instruction were rated as moderately aligned in terms of fidelity to the program model. Thus, in these eight schools where the ERO programs were designated as having reached a level of moderate alignment overall, at least seven out of up to 11 constructs included in the composites were rated as being moderately or well aligned according to the criteria presented in the observation protocols. These eight schools include four Reading Apprenticeship schools and four Xtreme Reading schools.

Schools identified as having especially problematic program implementation were those schools whose average fidelity rating on either the classroom learning environment dimension or

the comprehension instruction dimension was classified as implementation poorly aligned to the program models. The bottom panel of Table 3.5 also shows that 10 of the 34 high schools were found to have encountered serious implementation problems on at least one of the two key program dimensions during the first year of the study. Three of these schools demonstrated poorly aligned implementation on both the learning environment and the comprehension instruction dimensions; six demonstrated poorly aligned implementation only on the comprehension instruction dimension; and one demonstrated poorly aligned implementation only on the classroom learning environment dimension. These 10 high schools that encountered serious implementation problems include six Reading Apprenticeship schools and four Xtreme Reading schools.

Summary and First-Year Implementation Challenges

Both of the ERO programs were complex and multidimensional interventions being implemented by teachers who had no prior formal experience with supplemental reading instruction for adolescents. Each of the program developers provided a five-day summer training institute prior to the start of the first year of the study. During the school year, teachers attended two 2-day booster training sessions, and coaches from the developer teams made a minimum of three coaching visits to each teacher.

In all, the ERO programs in 24 of the 34 schools were found to have reached a level of implementation at least moderately aligned with the program models. The ERO programs in 16 of these schools were found to have reached a level of implementation well aligned with the models, indicating that almost all of the key implementation components were moderately aligned or well aligned with the characteristics of the program models. The implementation of the ERO programs in the remaining 10 schools was found to be especially problematic, and these programs were deemed to be poorly aligned reflections of their intended models.

ERO implementation in the 2005-2006 school year occurred in the context of three challenges that were distinctive to the first year of the project:

- The delayed start of the ERO classes in all schools
- The delayed acquisition of some prescribed program materials and resources
- The newness of the programs to the schools and the ERO teachers

As is discussed in Chapter 4, ERO classes began an average of six weeks after the start of the school year, and 16 of the participating schools started their ERO programs during the eighth week of school or later. As a result, more than two months had elapsed between the summer training institute and the start of the ERO classes. This caused disruptions in students' class schedules, and teachers were left an average of less than seven and a half months to try to

cover curricula intended for a nine-month school year. In response to this shorter time line, the developers were able to make some adjustments to compact their curricula. Nonetheless, teachers also were not able to get through all of the curricular units.

Each ERO classroom was intended to have the following components: a library, a file cabinet, a flipchart, an overhead projector, two computers, and a printer/scanner. These resources were to be purchased by the school district, using funds from its SLC grant. The ERO study team visited each of the participating schools within approximately four weeks of the start of the ERO classes and found that one or more of these classroom components were missing in 23 of the 34 schools. They communicated these findings to the district coordinators, reminding them of the expectation that the grant funds would be used to provide these components.

Until districts were able to provide the components, teachers made accommodations by borrowing overhead projectors or file cabinets, for example, until there were provided permanently. The most commonly missing items were computers. In these cases, the ERO teachers were advised by the program developers — whose staff were also making visits to sites and were aware of which teachers were missing materials — to postpone using the software programs they provided until the second semester. By the second semester of the year, all supplies had been provided to 27 of the 34 schools. The study team continued to communicate with the other seven high schools and their districts to encourage them to obtain the rest of their supplies.

All ERO teachers were new to the program they were trying to implement. They were learning the Reading Apprenticeship or Xtreme Reading program while teaching it, adding to the challenge of achieving high-fidelity implementation. In addition, three of the 34 teachers who attended the summer training institutes left their ERO teaching position before the end of the academic year.¹⁰ The schools that lost teachers had to conduct a search for replacements who met the eligibility criteria for the project (holding a high school teaching certificate in social studies or English and having at least a year of teaching experience).¹¹ These teachers were then trained in the relevant ERO program.

Each of these challenges was addressed systematically in the second year of the study. ERO classes began within an average of approximately two weeks of the start of the school year and started on the first day of school at 18 of the 34 schools. All the required equipment and supplies were provided to each of the ERO classrooms. Twenty-seven of the 34 teachers of the ERO classes at the end of the first year of implementation returned to teach the program again

¹⁰One of the three departing teachers left after having participated in the summer training but before the ERO course had started. The other two teachers left approximately half way through the school year.

¹¹The study team worked with the U.S. Department of Education officials responsible for the grant administration and the evaluation and the grantees to identify suitable replacement teachers and to schedule them for training and coaching.

in the second year.¹² All of the continuing and replacement teachers remained with the programs throughout the second year of the study. Thus, the second report from the study will provide information about both the implementation and impact of the ERO programs under conditions of a timelier start-up, better-equipped classrooms, and more experienced teachers than existed in the first year of implementation. In fact, results from classroom observations in the fall of the second year — the first of two second-year site visits — indicate that 31 of the 34 schools had reached at least a moderate level of alignment in terms of implementation on both of the key program dimensions and that 20 of the programs were well aligned with the program models on both implementation dimensions. Classroom observations conducted during the study's second year used the same protocols and process as those conducted in the first year of implementation, except that only one observer visited the classrooms rather than two.

¹²Twenty-five of these teachers taught the ERO courses the entire year. Two of the returning teachers replaced other ERO teachers in the middle of the first year, and thus returned the second year having taught the ERO course less than a full year.

Chapter 4

Student Attendance in the ERO Classes and Participation in Literacy Support Activities

In addition to examining the fidelity with which the sites participating in the Enhanced Reading Opportunities (ERO) study implemented the models of the two supplemental high school literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — the evaluation also includes an assessment of how much students participated in the ERO classes and whether they participated in other literacy support services either in or outside school. The evaluation team collected data about the frequency with which the ERO classes met and about whether and how often students attended. These data provide an indication of the overall “dosage” of the ERO interventions that students in the ERO group received during the first year of the study. The impact of the ERO programs will be a function, in part, of how much exposure the ERO students have to the classes throughout the school year. These data also provide an indication of whether students in the non-ERO group inadvertently enrolled in the ERO classes and thus diluted the overall contrast in literacy services received by students in the ERO and non-ERO groups.

The ERO evaluation team also collected data on the frequency with which students participated in classes or tutoring services that aimed to improve students’ reading and writing skills. Specifically, the student follow-up survey asked several questions about the frequency and duration with which students participated in such activities either in school or outside school. These data are available for students in both the ERO and the non-ERO groups and are intended to capture participation in both the ERO classes and other literacy support programs and services. They provide a measure of the difference in exposure to supplemental literacy support services between the ERO and non-ERO groups — which is a key factor in whether the ERO programs offer a contrast to the services that would otherwise be available.

This chapter discusses the following key findings:

- The ERO classes began an average of six weeks after the start of the school year and operated for an average of just over seven and a half months of the nine-month school year.
- More than 95 percent of the students in the ERO group enrolled in the ERO classes, and 91 percent were still attending the classes at the end of the school year.

- Students attended 83 percent of the scheduled ERO classes each month, and they received an average of just over 11 hours of ERO instruction per month.
- There were no systematic differences in ERO class enrollment and attendance rates between schools using Reading Apprenticeship and those using Xtreme Reading.
- Students who were randomly assigned to the study's ERO group reported a much higher frequency of participation in supplemental literacy services (in ERO classes and otherwise) than students in the non-ERO group. Although the largest difference occurred in a school-based literacy class, ERO students were also more likely to report working with a tutor in and outside school and attending a literacy class outside school.

In general, the ERO classes served as the primary source of literacy support services for students in the study sample. For students in the study's ERO group, the ERO classes substituted for a scheduled elective class — such as a career/technical education class, an arts class, a physical education or health class, or a foreign language class — and not one of the core-content classes: English/language arts, history/social studies, mathematics, and science. The ERO classes were not a source of literacy support for non-ERO students. Seven out of the 1,428 students in the non-ERO group enrolled in the ERO classes. Also, given that the ERO teacher at each school taught no other classes other than the ERO class, the only way for non-ERO students to receive ERO instruction was through enrollment in the ERO classes.

Student Enrollment and Attendance in the ERO Classes

The amount of ERO instruction that students receive is a function of program duration and student attendance. The longer the duration of the program, the greater the opportunity students have to participate in the ERO classes. The more often students attend, the more ERO instruction they will be exposed to. Following is an overview of findings from the evaluation's analysis of program duration and attendance.

Program Duration

The ERO programs were designed to operate for the full school year and to provide students with approximately nine months of supplemental literacy instruction. In fact, the ERO classes began an average of six weeks after the start of the 2005-2006 school year, ranging from

three to ten weeks across the 34 high schools.¹ The delayed start-up of the classes meant that the ERO programs operated for an average of just over seven and a half months rather than the full nine months of the school year.² This ranged from six and a half months in one school to eight and a half months in three schools. On average, across the participating high schools, students in Cohort 1 of the study sample had the potential to experience about 85 percent of the full planned Reading Apprenticeship and Xtreme Reading programs. Overall, during the first year of the project, 22 of the 34 participating high schools operated their ERO programs for more than seven and a half months.

Conducting the student recruitment and random assignment process at the start of the school year also meant that student class schedules had to be changed for the individuals assigned to the ERO group. This disrupted ERO students as they were pulled from elective classes and placed into ERO classes. In interviews with the study team, many of the ERO teachers reported that it took several days for students to settle in to their new schedules and adjust to the new expectations and routines.

Student Enrollment and Attendance

As part of their responsibilities to the project, the ERO teachers were required to maintain and report to the study team daily attendance records for all students randomly assigned to the ERO group. They were also asked to determine whether chronically absent students were still enrolled in the ERO programs or had transferred to another school in the district. These data, along with information about the length of ERO class periods, provided the basis for calculating several measures of ERO enrollment and attendance. These measures are displayed in Table 4.1.³

Overall, nearly 96 percent of students in the ERO group attended at least one ERO class during the year, and approximately 91 percent were still attending ERO classes at the end of the school year. On average, students remained enrolled in the ERO programs for just over seven months during the school year. Table 4.1 shows that similar percentages of students enrolled in and remained in the Reading Apprenticeship and the Xtreme Reading classes.

¹Because the selection of districts to receive the special SLC grants did not occur until June 2005, the student recruitment process was delayed until the start of the 2005-2006 school year. This required between three and 10 weeks to complete.

²Each of the participating high schools was in session for approximately nine months, excluding vacations.

³The findings presented in Table 4.1 are based on attendance data for ERO group students in the follow-up respondent sample — the same sample as is used in the impact analysis for this report. The ERO enrollment and attendance findings for these students provide an assessment of the dosage of ERO program services that is associated with the impact findings discussed in Chapter 5. Note that all measures in Table 4.1 include students from the ERO group who never enrolled in the ERO classes and students who left the program during the school year. Zero values were included for these students during the periods when they were not enrolled in the programs.

The Enhanced Reading Opportunities Study

Table 4.1

**Attendance in ERO Classes,
Follow-Up Respondent Sample in the ERO Group**

Characteristic	All Schools	Reading Apprenticeship Schools	Xtreme Reading Schools
Ever attended an ERO class during the year (%)	95.5	94.9	96.0
Attending ERO classes at the end of the year (%)	91.2	91.0	91.4
Average daily attendance rate in ERO classes per month ^a (%)	82.7	81.7	83.6
Number of months ERO program was in operation	7.7	7.8	7.7
Average number of months attending ERO classes	7.1	7.1	7.1
Average number of hours ERO class met per month	13.6	13.5	13.7
Average number of hours student attended ERO class per month	11.3	11.2	11.5
Sample size	1,408	686	722

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study monthly attendance data.

NOTES: Tests of statistical significance were not performed.

^aThere were 64 students who never attended an ERO class, 35 students from Reading Apprenticeship schools and 29 students from Xtreme Reading schools. Excluding these students, the average daily attendance rate for the remaining students who attended at least 1 ERO class is 86.6 percent for all schools, 86.1 percent for Reading Apprenticeship schools, and 87.1 percent for Xtreme Reading schools.

The ERO programs were designed for an average of 3 hours and 45 minutes of class time per week (which is scheduled either as 45-minute classes each day or as 80- to 90-minute classes every other day). With an average of 20 days of school per month, the ERO classes were designed to provide students with approximately 15 hours of supplemental literacy instruction per month. Based on the attendance data provided by the ERO teachers, Table 4.1 shows that the ERO classes met for an average of 13.6 hours per month (approximately 3 hours and 25 minutes per week). On average, students in the ERO group attended 82.7 percent of the scheduled ERO classes each month. This amounts to an average of 11.3 hours of ERO instruction per month, or just under 3 hours per week.

Student Participation in Literacy Support Activities

A requirement of the ERO funding grants from ED was that the participating schools would not operate other supplemental literacy programs during the evaluation period. This was to ensure that the effectiveness of the ERO programs could be evaluated in a context where they were not being compared with similar interventions. School district officials were asked in their grant applications to affirm that none of the schools included in their grant applications were currently using or planning to implement supplemental adolescent literacy programs for their ninth-grade students.⁴ At the same time, students in both the ERO and the non-ERO group were free to seek out other literacy-related services on their own. In some cases, they found other adults in the school to provide tutoring; in other cases, students and their families sought out other classes or tutors outside school.

This section of the chapter examines the extent to which the availability of the ERO programs created a sharp contrast in ERO students' exposure to supplemental literacy services in and outside school compared to students in the study's non-ERO group. To the degree that students in the non-ERO group participated in supplemental literacy support services either in or outside school, the overall contrast with the ERO group's participation in the ERO classes would be reduced. Before turning to this analysis, the chapter first reviews the manner in which the ERO classes were inserted into students' course schedules and discusses the degree to which literacy instruction was embedded in the typical English/language arts classes in the participating high schools.

Elective Courses

The ERO class was intended to substitute for an elective class, rather than for a core academic class, in students' ninth-grade course schedules. Each of the participating high schools used scheduling models that allowed students to take seven or eight courses during the year. Four of these courses were academic requirements such as English/language arts (ELA), mathematics, history, and science, leaving three or four slots for elective classes. Even in high schools where one of those slots was filled with another required course like physical education or health, there were still two or three slots open for electives. Thus, the primary difference between the ERO group and the non-ERO group is that the ERO students had one of their elective classes replaced by the ERO class and that the non-ERO students remained in their elective classes. This section of the chapter discusses the nature of the classes taken by the non-ERO group. These constitute a primary feature of the "counterfactual" to the ERO classes.

⁴U.S. Department of Education (2005).

A review of class schedules for students in the ERO group confirmed that the ERO classes did not replace English/language arts or other required academic classes (mathematics, history, or science). All students in the ERO and non-ERO groups were enrolled in the academic classes that were required by the school or district. The class schedules for the non-ERO group show that alternatives to the ERO classes consisted of a wide array of ninth-grade elective classes, over three-quarters of which fell into four main categories. Twenty-five percent of these classes were in the subject area of career and technical education; 21 percent were in the visual and performing arts; 16 percent were in physical education/health; 15 percent were in a foreign language.⁵ That is, with few exceptions, students in the non-ERO group were *not* enrolled in the ERO classes and, instead, were enrolled in a variety of electives.⁶ However, because ERO students had room for one or more electives beyond the ERO class in their schedules, this same variety of elective classes included students from both the ERO and the non-ERO group. In these shared courses, though, students in the ERO group were underrepresented relative to students in the non-ERO group. In short, non-ERO students were typically enrolled in four or five core required classes and two or three elective classes, while ERO students were typically enrolled in the same four or five core required classes, one or two of the same elective classes, and the ERO class. The ERO class never substituted for one specific elective for all ERO students at any given school.

To demonstrate how ERO fits into student schedules, two examples are presented in Table 4.2. Between them, these examples represent the three most common types of variation in student schedules: the schedule model, the number of course slots within the schedule model, and the number of required courses. First, the two most commonly used schedule models in the 34 high schools were the *traditional bell schedule*, in which each class typically meets daily for 40 to 50 minutes (Example 1); and the *alternating (or A/B) block schedule*, in which each class meets for about 90 minutes every other day (Example 2). Second, since the modal number of course slots in the schools' schedule models was 8.0 and the mean was 7.7, one example reflects a schedule with seven course slots, and the other has eight course slots. Lastly, as noted above, some schools may have included another required course (for example, physical education or health) beyond the four core academic courses.

Interviews with elective teachers supplemented the data about elective courses that were obtained from student schedules. Specifically, these interviews provided data about whether the elective courses focused explicitly on teaching reading and writing skills, thus offering students a similar opportunity as the ERO classes. For the few non-ERO elective classes

⁵These figures are based on a more detailed analysis of the 319 elective courses listed on student schedules from 10 of the 34 ERO high schools, one school from each district.

⁶Seven out of 1,428 non-ERO students from the study sample were found to have enrolled in an ERO class.

The Enhanced Reading Opportunities Study

Table 4.2

Comparison of ERO and Non-ERO Student Schedules

Example 1: Traditional Bell Schedule, Seven Periods, Four Required Courses

Period	ERO Students	Non-ERO Students
1	English/Language Arts	English/Language Arts
2	Math	Math
3	Science	Science
4	Social Studies/History	Social Studies/History
5	<i>ERO</i>	<i>Elective</i>
6	Elective	Elective
7	Elective	Elective

Example 2: Alternating (A/B) Block Schedule, Eight Periods, Five Required Courses

Period	ERO Students		Non-ERO Students	
	Day A	Day B	Day A	Day B
1	English/Language Arts	Science	English/Language Arts	Science
2	Math	Social Studies/History	Math	Social Studies/History
3	Required course	<i>ERO</i>	Required Course	<i>Elective</i>
4	Elective	Elective	Elective	Elective

NOTE: These are not actual schedules but represent two types of schedules in ERO high schools. They are used to demonstrate how ERO fits into student schedules.

where reading and writing were taught more explicitly, there were both ERO and non-ERO students indicating that exposure to these types of literacy supports was distributed across both groups. That is, the enrollments in these courses did not exclusively represent one group or the other, nor were all students from one group or the other enrolled in these courses. Specifically, in three different high schools, three courses were identified that had explicit literacy instruction, and they enrolled an average of 10 non-ERO students and five ERO students. These are the only three of hundreds of elective courses taken by students in the non-ERO and ERO groups across the 34 high schools in the study that were judged to include explicit literacy instruction. Even here, the classes enrolled a small proportion of the non-ERO group, and they included similar numbers of ERO and non-ERO students.

English/Language Arts Instruction

ELA classes offered another venue where literacy instruction might occur beyond elective courses and different kinds of supplemental literacy services. Both ERO and non-ERO students were enrolled in ELA classes together, and they received the same amount of ELA instruc-

tion. Interviews were conducted with ELA teachers that investigated the nature of the ELA instruction at the 34 high schools, with particular focus on assessing whether literacy-rich ELA instruction was already occurring. Because the ELA instruction was the same for ERO and non-ERO students, literacy-rich ELA instruction would not cause differences between those two groups of students but would, rather, possibly decrease the potential value added by the ERO classes. In interviews with members of the study team, ELA teachers across all of the participating schools indicated that their classes were comprised primarily of exposing students to different literary genres and some instruction in grammar and composition. While there was regular use of reading and writing activities, the instruction was literature-based and was not focused explicitly on improving reading and writing skills with the intensity or specificity of the ERO classes.

Overall, the support for building students' literacy skills available in the ninth-grade year to students in the non-ERO group through ELA and elective classes was not comparable in focus and intensity to that provided by the ERO classes. The ERO classes offered a strong contrast to the experiences of the non-ERO students, and they were different from other elective classes in their focus on literacy instruction. While the ERO programs were not taught in a literacy vacuum (that is, all students had reading and writing activities as part of their courses), they did provide support to students that was different and more intensive than what they typically received.

Student Participation in Supplemental Literacy Support Activities

The student follow-up survey included items aimed at determining the amount of extra literacy support that students received during the school year, beyond their regular English/language arts class. The survey asked about four categories of extra literacy help: classes in school, classes outside school, an adult tutor in school, and an adult tutor outside school. The first category describes such supports as the ERO courses. This item essentially provides an opportunity for ERO students to report on their attendance in the ERO classes, and for non-ERO students to report on their participation in literacy support activities that would be most similar to or “competitive” with ERO. The other three categories of activities cover other ways in which students might receive help with their reading and writing skills.

The survey questions asked all students about how long (duration) and how often (frequency) they participated in each of the four categories of activities. For example, a student who attended a “help” session every day for the full school year was projected to have attended approximately 180 sessions (about 20 days per month for nine months, or the typical number of days in a school year). Similarly, a student who reported attending twice per week for a semester was projected to have attended about 36 sessions (eight days per month for about four and a half months).

Table 4.3 provides the average levels of student participation in these four types of supplemental literacy support activities. The table also includes estimates of the differences in participation between the ERO and non-ERO groups.

The comparisons of the two groups provide an indication of the increase in literacy instruction and support that the ERO programs produce over and above what students would be exposed to without the programs. Reflecting their participation in the ERO program, students in the ERO group participated in a school-based literacy class six times more than students in the non-ERO group. It should be noted, however, that students in the non-ERO group did report receiving some exposure to a literacy class in school, though only a handful of non-ERO students across all the high schools ever enrolled in an ERO class. Table 4.3 also shows that students in the ERO group also reported higher levels of participation in tutoring sessions and in literacy classes outside school than students in the non-ERO group.

The Enhanced Reading Opportunities Study

Table 4.3

**Participation in Supplemental Literacy Support Activities,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	Non-ERO Group	Impact	Impact Effect Size	P-Value for the Difference
<u>All schools (number of sessions)</u>					
School-based literacy class	63.2	11.4	51.8 *	1.44 *	0.000
School-based adult tutor	22.0	8.2	13.9 *	0.46 *	0.000
Outside-school literacy class	5.5	2.5	3.0 *	0.20 *	0.001
Outside-school adult tutor	8.6	5.5	3.1 *	0.13 *	0.011
Sample size	1,410	1,002			
<u>Reading Apprenticeship schools (number of sessions)</u>					
School-based literacy class	64.1	11.6	52.5 *	1.46 *	0.000
School-based adult tutor	21.0	8.5	12.6 *	0.42 *	0.000
Outside-school literacy class	5.1	3.6	1.5	0.10	0.302
Outside-school adult tutor	8.8	7.1	1.7	0.07	0.356
Sample size	689	455			
<u>Xtreme Reading schools (number of sessions)</u>					
School-based literacy class	62.3	11.0	51.3 *	1.43 *	0.000
School-based adult tutor	23.0	8.0	14.9 *	0.49 *	0.000
Outside-school literacy class	5.8	1.4	4.4 *	0.29 *	0.000
Outside-school adult tutor	8.5	4.1	4.3 *	0.19 *	0.007
Sample size	721	547			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-based class standard deviation = 35.924; school-based tutor standard deviation = 30.240; outside-school class standard deviation = 14.896; outside-school tutor standard deviation = 23.027).

A two-tailed t-test was applied to the impact estimate. The statistical significance level is indicated (*) when the p-value is less than or equal to 5 percent.

For each of the above measures, data are missing for no more than 6 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

Chapter 5

Early Impacts on Student Reading Achievement and Reading Behaviors

The primary focus of the Enhanced Reading Opportunities (ERO) evaluation is to assess the impact of supplemental literacy interventions on adolescent students' reading comprehension skills and behaviors and on their overall academic performance during high school. The early impact analysis presented in this report addresses two questions that pertain to the first year in which the ERO programs were being implemented and to their effects for ninth-grade students at the end of the year in which they were enrolled in the programs:¹

- What is the impact of supplemental literacy programs on ninth-grade students' reading comprehension as measured by standardized test scores for reading comprehension and reading vocabulary?
- What is the impact of supplemental literacy programs on ninth-grade students' vocabulary and on their reading behaviors as measured by self-reported information about how much students read and whether they use specific reflective reading strategies?

Because the study's two supplemental literacy programs — Reading Apprenticeship Academic Literacy and Xtreme Reading — focus on producing immediate improvements in students' reading comprehension ability, the early impact analysis presented in this report places a higher priority on the first question above. Each of the programs also endeavors to enhance students' vocabulary and their interest in reading both in and outside school and to increase their use of strategies that are characteristic of proficient readers. For this reason, the analysis also examines impacts on vocabulary test scores and on three measures of students' reading behaviors. As discussed in Chapter 2, measures of students' reading comprehension and vocabulary skills are drawn from their performance on the Group Reading Assessment and Diagnostic Examination (GRADE) administered at the end of their ninth-grade year. The measures of reading behavior were developed from the follow-up survey that was administered to students in the study sample at the end of their ninth-grade year.

This chapter first presents early impact findings for all 34 of the high schools in the evaluation. The results that are pooled across the two programs selected for the demonstration

¹Subsequent reports will also examine impacts on a range of longer-term outcomes, including performance on standardized state tests, credits earned toward graduation, daily attendance, grade-to-grade promotion rates, and dropout rates.

provide evidence about the effectiveness of the two supplemental literacy interventions selected by the expert panel for this project as a class of interventions. The chapter then presents findings for each of the two ERO programs separately. Although Reading Apprenticeship and Xtreme Reading share overarching goals for adolescent literacy development and share many instructional principles, these results provide evidence about whether their differences in operating strategies resulted in different patterns of impacts.

The chapter also summarizes findings for subgroups of students defined by pre-random assignment background characteristics, including their baseline reading test scores, whether they had repeated an earlier grade, and whether a language other than English is spoken at home.

The chapter ends with an exploration of variation in impacts across two subgroups of schools in the study. The implementation of the ERO programs in one group of schools was classified as at least moderately aligned with the program models — as defined in Chapter 3 — and the schools were able to operate their ERO programs for more than seven and a half months (the average for the sample as a whole). The implementation of the ERO programs in the other group of schools were classified as poorly aligned with their program models or they operated for seven and a half months or less. It is not possible to conclude definitively that differences in impacts between these two groups of schools were *caused* by differences in their early start-up experiences. Rather, this analysis represents an exploration of impacts under conditions that were more like those intended by the program developers.

The chapter discusses the following key findings:

- Overall, the ERO programs produced a positive and statistically significant impact on reading comprehension test scores, with an effect size of 0.09 standard deviation. This impact corresponds to an improvement from the 23rd percentile nationally, as represented by the average scores for students in the non-ERO group, to the 25th percentile nationally, as represented by the average scores for students in the ERO group.
- Despite the positive impact on reading comprehension test scores, almost 90 percent of students in the study sample who enrolled in the ERO programs were still reading below grade level at the end of the ninth grade.
- Although they are not statistically significant, the magnitudes of the impact estimates on reading comprehension test scores for each literacy intervention are the same as those for the full study sample.
- The ERO programs did not produce statistically significant impacts on vocabulary test scores.

- The ERO programs exhibited a mix of positive and negative impacts on the measures of reading behavior, but these are not statistically significant.
- Positive impacts on reading comprehension were concentrated among schools whose implementation of the ERO programs was at least moderately aligned with the program models and schools that were able to operate their ERO programs for more than seven and a half months.

Early Impacts on Reading Achievement

The ERO study assesses the impact of supplemental literacy interventions of the type represented by Reading Apprenticeship and Xtreme Reading. As such, the analysis focuses first on impacts that are pooled across both interventions and all sites in the study sample. Thus, in pooling the sample across all schools in the study, the analysis has sufficient power to detect statistically significant impacts that are somewhat smaller than those that can be detected for each ERO program separately. At the same time, the study was designed to ensure adequate statistical power for policy-relevant impact estimates from each intervention separately. The primary measure of reading achievement for this study is students' scores on the GRADE reading comprehension assessment. A secondary measure of students' reading achievement is their scores on the GRADE vocabulary assessment.

- **Overall, the ERO programs produced a positive and statistically significant impact on reading comprehension (0.90 standard score point, which corresponds to an effect size of 0.09 standard deviation).**

The first row in Table 5.1 shows that, averaged across all 34 participating high schools, the ERO programs improved reading comprehension test scores by 0.9 standard score point and that this impact is statistically significant (p-value is less than or equal to 5 percent). Expressed as a proportion of the overall variability of test scores for students in the non-ERO group, this represents an effect size of 0.09 (or 9 percent of the standard deviation of the non-ERO group's test scores). Table 5.1 also shows that this impact corresponds to an improvement from the 23rd percentile nationally, as represented by the average scores for students in the non-ERO group, to the 25th percentile nationally, as represented by the average scores for students in the ERO group.

Figure 5.1 places this impact estimate in the context of the actual and expected change in the ERO students' reading comprehension test scores from the beginning of ninth grade to the end of ninth grade. The bottom section of the bar shows the average reading comprehension test score for students in the ERO group at the beginning of their ninth-grade year. This average of 85.9 standard score points corresponds, approximately, to a grade equivalent of 5.1 and indicates an average reading level for students nationally at the start of fifth grade. This marks the

The Enhanced Reading Opportunities Study

Table 5.1

**Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Reading comprehension					
Average standard score	90.1	89.2	0.9 *	0.09 *	0.019
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	25	23			
Reading vocabulary					
Average standard score	93.4	93.2	0.3	0.03	0.472
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	32	31			
Sample size	1,408	1,005			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment.

NOTES: The follow-up GRADE assessment was administered in the spring of 2006 near the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form B). No statistical tests or arithmetic operations were performed on these reference points.

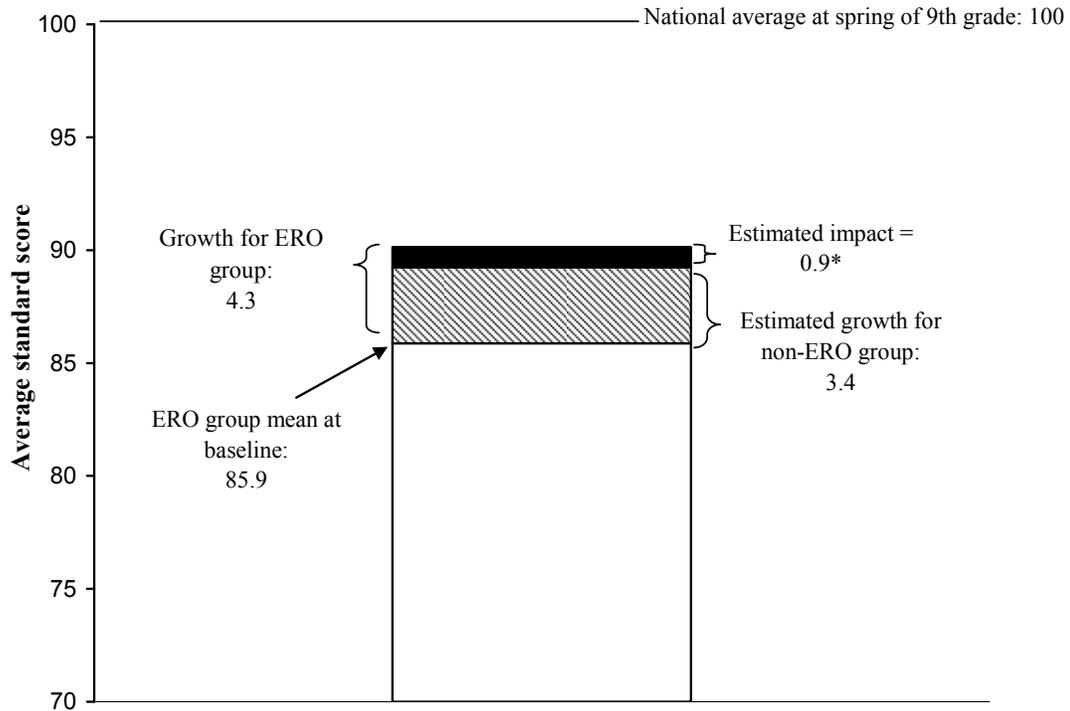
The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

The Enhanced Reading Opportunities Study

Figure 5.1
Impacts of Reading Comprehension,
Cohort 1 Follow-Up Respondent Sample



SOURCES: MDRC calculations from the Enhanced Reading Opportunities Study baseline and follow-up GRADE assessments; American Guidance Service, *Group Reading Assessment and Diagnostic Evaluation: Teacher's Scoring and Interpretive Manual, Level H*.

NOTES: The baseline GRADE assessment was administered in the fall of 2005 at the start of students' ninth grade year and prior to their random assignment to the ERO and non-ERO groups. The follow-up GRADE assessment was administered in the spring of 2006 near the end of students' ninth-grade year.

The ERO group growth at follow-up is calculated as the difference between the unadjusted ERO group mean at baseline and the unadjusted ERO group mean at follow-up. The impact was estimated using ordinary least squares and adjusted to account for the blocking of random assignment by school and to control for random differences between the ERO and non-ERO groups in baseline reading comprehension test scores and age at random assignment. The expected ERO group growth at follow-up is the difference between the actual ERO group growth and the impact.

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

The national average for standard score values is 100, and its standard deviation is 15.

Rounding may cause slight discrepancies in calculating sums and differences.

starting point for measuring both the observed growth in their reading achievement through the end of their ninth-grade year and their expected growth to be estimated through the test scores of the non-ERO group at the end of ninth grade.

Together, the bottom two sections of the bar in Figure 5.1 show the estimated reading comprehension test scores of students in the non-ERO group at the end of their ninth-grade year. The middle section of the bar, therefore, represents the growth in test scores experienced by the non-ERO group. This growth of 3.4 points provides the best indication of what the ERO group would have achieved during their ninth-grade year had they not had the opportunity to attend the ERO classes. The top section of the bar shows the ERO impact on reading comprehension test scores. Thus, the impact of the ERO programs represents a 26 percent improvement over and above what the ERO group would have achieved if they had not had the opportunity to attend the ERO classes.² From this perspective, the ERO programs produced more progress on reading comprehension than the gains expected for this sample of students had they not been selected for the programs.

Together, the top two sections of the bar in Figure 5.1 indicate that students in the ERO group improved by an average of 4.3 standard score points over the course of their ninth-grade year. Thus, the impact of the ERO programs accounts for 21 percent of the average test score improvement experienced by the ERO group.³

The solid line at the top of Figure 5.1 shows the national average (100 standard scale points) for students at the end of ninth grade, in the spring. Students scoring at this level are considered to be reading at grade level. Despite the program impact, therefore, students' reading comprehension scores still lagged nearly 10 points below the national average for performance on GRADE reading comprehension for students at the end of their ninth-grade year. In fact, almost 90 percent of the students in the ERO group had reading comprehension scores that were below grade level, and 76 percent had scores that were two or more years below grade level.

- **Although the difference is not statistically significant, vocabulary test scores for students in the ERO group were estimated to be 0.3 standard score point higher than those for the non-ERO group.⁴**
- **Estimated impacts on reading comprehension and vocabulary test scores for each ERO program are not statistically significant.**

²This was calculated by dividing the impact (0.9 standard score point) by the average improvement of the non-ERO group (3.4 standard score points).

³This was calculated by dividing the impact (0.9 standard score point) by the average improvement of the ERO group (4.3 standard score points).

⁴The ERO study did not include a vocabulary test at baseline. As a result, it is not possible to place the impacts on vocabulary in the context of changes that occurred over the course of students' ninth-grade year.

Table 5.2 shows that the impacts on reading comprehension for both Reading Apprenticeship and Xtreme Reading are of the same magnitude as that found for the full sample of schools in the study. However, neither of these results is statistically significant. The table also shows that neither ERO program produced a statistically significant impact on vocabulary test scores.

Early Impacts on Students' Reading Behaviors

As noted in Chapter 2, the Enhanced Reading Opportunities Student Follow-Up Survey was administered at the same time as the follow-up GRADE assessment, at the end of the students' ninth-grade year. The impact analysis presented in this chapter focuses on three measures of students' reading behavior that were derived from the survey: amount of school-related reading, amount of non-school-related reading, and use of reflective reading strategies.⁵ Table 5.3 presents early findings on the ERO programs' average impact on these three measures. Table 5.4 presents these results separately for each of the two ERO programs.

- **Overall, the ERO program impacts on the reading behavior measures were not statistically significant.**

Each of the two supplemental literacy programs seeks to motivate students to read more. They do this both by providing opportunities to read and discuss what they read in the ERO classes and by providing classroom libraries and assigning texts for students to read at home. The goal is to expose students to a wide range of reading opportunities, while building the strategies that proficient readers use and thereby stimulating students' interest in reading more both for school and for their own enjoyment.

Table 5.3 shows that, across all 34 high schools, the amount of reading that students in the ERO group reported is greater than that of students in the non-ERO group. Neither of these results is statistically significant. The impact on students' reports of using reflective reading strategies is nearly zero.

Table 5.4 shows the impacts on reading behaviors separately for each ERO program. Although the bottom panel of Table 5.4 indicates that Xtreme Reading produced a positive and statistically significant impact on the amount of school-related reading that students reported, this result should be interpreted cautiously. As noted in Chapter 2, the analyses include qualifying statistical tests aimed at assessing the robustness of multiple impacts within the reading behavior measurement domain. The qualifying tests examine the estimated impact on a composite index of reading behaviors for each ERO program separately and a test of whether the difference in

⁵A list of the survey items used to create these three measures is presented in Appendix A.

The Enhanced Reading Opportunities Study

Table 5.2

**Impacts on Reading Achievement,
Cohort 1 Follow-Up Respondent Sample,
by Program**

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Reading Apprenticeship schools</u>					
Reading comprehension					
Average standard score	89.8	88.9	0.9	0.09	0.097
<i>Corresponding grade equivalent</i>	6.1	5.9			
<i>Corresponding percentile</i>	24	23			
Reading vocabulary					
Average standard score	93.2	92.8	0.5	0.05	0.393
<i>Corresponding grade equivalent</i>	7.7	7.7			
<i>Corresponding percentile</i>	31	31			
Sample size	686	454			
<u>Xtreme Reading schools</u>					
Reading comprehension					
Average standard score	90.5	89.6	0.9	0.09	0.090
<i>Corresponding grade equivalent</i>	6.2	6.0			
<i>Corresponding percentile</i>	25	24			
Reading vocabulary					
Average standard score	93.6	93.5	0.1	0.01	0.846
<i>Corresponding grade equivalent</i>	7.8	7.8			
<i>Corresponding percentile</i>	32	32			
Sample size	722	551			
<hr/>					
Difference in Impacts Between Programs			Difference in Impacts	Difference in Impact Sizes	P-Value for Difference
<u>Reading Apprenticeship minus Xtreme Reading</u>					
Reading comprehension standard score			0.0	0.00	0.962
Reading vocabulary standard score			0.4	0.04	0.664

(continued)

Table 5.2 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment.

NOTES: The follow-up GRADE assessment was administered in the spring of 2006 near the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The national average for standard score values is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the *GRADE Teacher's Scoring and Interpretive Manual* (Level H, Grade 9, Spring Testing, Form B). No statistical tests or arithmetic operations were performed on these reference points.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

impacts between the two groups of schools is statistically significant.⁶ These tests indicate that neither ERO program produced a statistically significant impact on the composite index that was created to capture the three reading behavior measures. Also, the difference in the impacts on the composite index between the two programs was not statistically significant. As a result, the one statistically significant result presented in Table 5.4 should be interpreted cautiously.

Early Impacts for Subgroups of Students

While all students in the study sample had baseline reading comprehension skills between the fourth- through seventh-grade level at the start of ninth grade, the ERO study sample includes a diverse population of students. With this diversity in mind, the ERO evaluation was designed to allow for the estimation of impacts for key subgroups of students who face especially challenging barriers to literacy development and overall performance in high school. For example, prior research has shown that especially low literacy levels, evidence of failure in prior grades, and having English as a second language are powerful predictors of school success.⁷

This section of the chapter and Appendix H examine variation in ERO program impacts for subgroups of students defined by their baseline reading comprehension test scores, whether

⁶See Appendix E, Appendix Table E.3, for the results of these qualifying tests.

⁷Roderick (1993); Fine (1988).

The Enhanced Reading Opportunities Study

Table 5.3

**Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample**

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>All schools</u>					
Amount of school-related reading (prior month occurrences)	44.2	43.4	0.8	0.02	0.669
Amount of non-school-related reading (prior month occurrences)	27.3	26.0	1.3	0.04	0.315
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	-0.01	0.849
Sample size	1,410	1,002			

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

they were overage for the ninth grade, and whether a language other than English was spoken in their homes. As reported in Chapter 2 (see Table 2.4), 36 percent of the study sample had baseline test scores that indicated reading levels that were four to five years below grade level at the start of ninth grade, and another 28 percent were reading from three to four years below grade level. Also, over a quarter of the students in the study sample were overage for the ninth grade (that is, they were age 15 years or older at the start of ninth grade), which is used to indicate that

The Enhanced Reading Opportunities Study
Table 5.4
Impacts on Reading Behaviors,
Cohort 1 Follow-Up Respondent Sample,
by Program

Outcome	ERO Group	Non-ERO Group	Estimated Impact	Estimated Impact Effect Size	P-Value for Estimated Impact
<u>Reading Apprenticeship schools</u>					
Amount of school-related reading (prior month occurrences)	43.8	48.3	-4.5	-0.10	0.100
Amount of non-school-related reading (prior month occurrences)	26.8	27.6	-0.8	-0.02	0.672
Use of reflective reading strategies (4-point scale)	2.6	2.7	0.0	-0.03	0.600
Sample size	689	455			
<u>Xtreme Reading schools</u>					
Amount of school-related reading (prior month occurrences)	44.5	39.2	5.3 *	0.12 *	0.029
Amount of non-school-related reading (prior month occurrences)	27.7	24.6	3.1	0.10	0.081
Use of reflective reading strategies (4-point scale)	2.6	2.6	0.0	0.02	0.779
Sample size	721	547			
<u>Difference in Impacts Between Programs</u>					
			Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
<u>Reading Apprenticeship minus Xtreme Reading</u>					
Amount of school-related reading			-9.8 *	-0.22 *	0.007
Amount of non-school-related reading			-3.9	-0.12	0.133
Use of reflective reading strategies			0.0	-0.05	0.566

(continued)

Table 5.4 (continued)

SOURCE: MDRC calculations from the Enhanced Reading Opportunities follow-up student survey.

NOTES: The student follow-up survey was administered in spring 2006 at the end of students' ninth-grade year.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment. The ERO group value is the unadjusted mean for the students randomly assigned to the ERO programs. The non-ERO group value is calculated as the difference between the ERO group value and the estimated impact.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (school-related reading standard deviation = 43.867; non-school-related reading standard deviation = 31.834; use of reading strategies standard deviation = 0.670).

A two-tailed t-test was applied to the impact estimate and to the difference in impacts. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

For each of the above measures, data are missing for no more than 5 percent of the respondents.

Rounding may cause slight discrepancies in calculating sums and differences.

a student was retained in a prior grade.⁸ Approximately 45 percent of the students in the sample lived in households where a language other than English was spoken.

Table 5.5 provides a summary of impact findings for the subgroups of students defined by their baseline reading comprehension test scores, whether they were overage for the ninth grade, and whether a language other than English was spoken in their homes.⁹ In general, the table indicates that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores for two of the subgroups and on vocabulary test scores for one of the subgroups. Nevertheless, the composite qualifying statistical test for the multiple hypothesis tests reflected in the table indicates that the overall variation in impacts across the subgroups is not statistically significant (F-statistic = 0.865; p-value = 0.534). Also, the difference in impacts between subgroups was not statistically significant.¹⁰

The first column in Table 5.5 shows the impact on reading comprehension test scores in effect size units. It indicates that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores for students who were overage for grade and for students from multilingual families. However, the difference between these impacts and those for their counterpart subgroups of students are not statistically significant. As a result, although the

⁸National Center for Education Statistics (1990).

⁹Appendix Tables H.1 through H.6 in Appendix H provide the outcome levels for the ERO and non-ERO groups, the estimated impacts, impact effect sizes, and p-values for the estimates presented in Table 5.5. The tables in Appendix H also show the difference in estimated impacts across subgroups and p-values of these differences.

¹⁰See Appendix H. Also, as noted in Chapter 2, 423 students with baseline reading test scores that were not within the target range intended for the study were not included in the impact analysis for this report. Sensitivity tests of the impact estimates indicate that the findings are not sensitive to the inclusion of these students.

The Enhanced Reading Opportunities Study
Table 5.5

Impact Effect Sizes for Student Subgroups

	Number of Students	Reading Comprehension Impact		Vocabulary Impact		Amount of School-Related Reading Impact		Amount of Non-School-Related Reading Impact		Use of Reflective Reading Strategies Impact	
		Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value	Effect Size	P-Value
<u>Baseline reading comprehension</u>											
6.0-7.0 grade equivalent	855	0.10	0.107	0.12 *	0.040	0.02	0.760	0.11	0.126	-0.06	0.376
5.0-5.9 grade equivalent	680	0.08	0.274	-0.06	0.401	0.06	0.430	0.05	0.526	0.05	0.471
4.0-4.9 grade equivalent	878	0.08	0.233	-0.02	0.729	0.00	0.998	-0.03	0.691	0.00	0.956
<u>Overlap for grade^a</u>											
Students overage for grade	644	0.19 *	0.007	0.09	0.221	0.04	0.667	0.10	0.253	-0.03	0.676
Students not overage for grade	1,769	0.05	0.267	0.00	0.992	0.02	0.718	0.02	0.647	-0.01	0.876
<u>Language spoken at home</u>											
Students from multilingual families	1,133	0.12 *	0.027	0.10	0.072	0.12	0.052	0.12 *	0.031	-0.03	0.664
Students from English-only families	1,280	0.07	0.181	-0.03	0.512	-0.09	0.140	-0.05	0.387	-0.01	0.908

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment and follow-up student survey.

NOTES: Appendix H provides detailed information about each of the student subgroup impact estimates, including outcome levels for the ERO and non-ERO groups, impact estimates, p-values, and differences in impacts among subgroups.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505; school-related reading = 43.867; non-school related reading = 31.834; use of reading strategies = 0.670).

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

^aA student is defined as overage for grade if he or she turned 15 before the start of ninth grade.

ERO programs produced a statistically significant impact on reading comprehension test scores for these subgroups of students, the analysis does not provide adequate confidence to conclude that the programs “worked better” for those students than it did for other subgroups of students.

The second column of Table 5.5 shows that the ERO programs produced a positive and statistically significant impact for students with baseline reading comprehension test scores that fell between the 6.0 and 7.0 grade equivalent. The difference between this impact and those for the other two test score subgroups is not statistically significant.

The far-right columns of Table 5.5 summarize the impacts on the reading measures for each of the subgroups. They indicate that the ERO programs produced a positive and statistically significant impact on the amount of non-school-related reading reported by students from multilingual families. In addition, the difference in impacts on school-related reading between students in multilingual families and students from English-only families is statistically significant. The qualifying tests that were conducted to account for the multiple hypothesis tests, however, indicate that the ERO program impacts on the composite index that was created to capture the three reading behavior measures are not statistically significant. Thus, the single statistically significant impact on reading behaviors in Table 5.5 should be interpreted cautiously.

The Relationship Between Early Impacts and First-Year Implementation Issues

This section of the chapter explores the variation in impacts of the ERO programs under conditions that were more or less proximal to those intended at the outset of the study and, as noted in Chapter 3, are more prevalent in the study’s second year than they were in the study’s first year. Specifically, it examines impacts for subgroups of the participating high schools that were defined by the degree to which they were able to achieve two implementation milestones during the first year of the study: whether they reached at least a moderate level of implementation fidelity (as defined in Chapter 3) and whether they were able to operate for more than seven and a half months (the average for the sample). The 15 schools that were able to reach both of these thresholds were deemed to have had a first-year start-up experience that was more in line with the original intent of the project than schools that did not meet these thresholds.

It is important to note that the analyses presented in this section of the chapter are exploratory and are not able to establish causal links between these early implementation milestones and variation in estimated impacts on student’s reading achievement across the sites. A variety of other program and school characteristics — not examined in the analyses presented here — may also be associated with differences in impacts across the schools. As an exploratory analysis, it is also not appropriate to extrapolate from these findings to predict the impact of the ERO programs in the second year of the project.

The exploration of relationships between impacts and first-year implementation challenges proceeds in three stages. The first stage provides an assessment of overall variation in impacts on reading comprehension test scores across the 34 participating schools. To the degree that there is variation in impacts across the sites, the overall average may be masking important differences in the effectiveness (or lack of effectiveness) of the ERO programs under some conditions. The second stage explores two sets of relationships: (1) the relationship between impacts and the implementation fidelity ratings and (2) the relationship between impacts and program implementation duration. The third stage combines the two indicators of first-year implementation challenges and presents impacts for two groups of sites based on whether they encountered serious problems either with implementation fidelity or with program duration during the first year of the study.

Overall Variation in ERO Impacts across Schools

Figure 5.2 illustrates the variation in estimated program impacts on reading comprehension scores across the 34 participating high schools.¹¹ For each school and for the overall average, the figure displays mean impact estimates (represented by the squares) and the 95 percent confidence intervals around the mean impact estimates (represented by the lines extending above and below the squares.) Here, the wider the confidence interval, the broader the margin of error and the greater the uncertainty about the impact estimate. Confidence intervals that do not include zero are statistically significant (p-value is less than or equal to 5 percent). The school-by-school impact estimates range from an ERO program producing a reduction in reading comprehension test scores of 7.1 standard score points to an ERO program producing an increase of 5.9 standard score points. In all, 23 estimates are positive, and 11 are negative; 16 estimates are smaller than the full-sample average, and 18 estimates are about the same or larger. Only five of the school-level impact estimates are statistically significant.

The variation in estimated impacts displayed in Figure 5.2 overstates the variation in true impacts, however, because a large portion of the variation in estimated impacts is due to estimation error. In other words, many of the estimates in the figure appear to be highly negative or highly positive; yet, for all but five of the estimates, their confidence intervals include zero, which indicates that they cannot be distinguished reliably from zero. For example, the second-most-negative impact is -3.7 standard score points, but its confidence interval ranges from -7.8 to 0.4 standard score points.

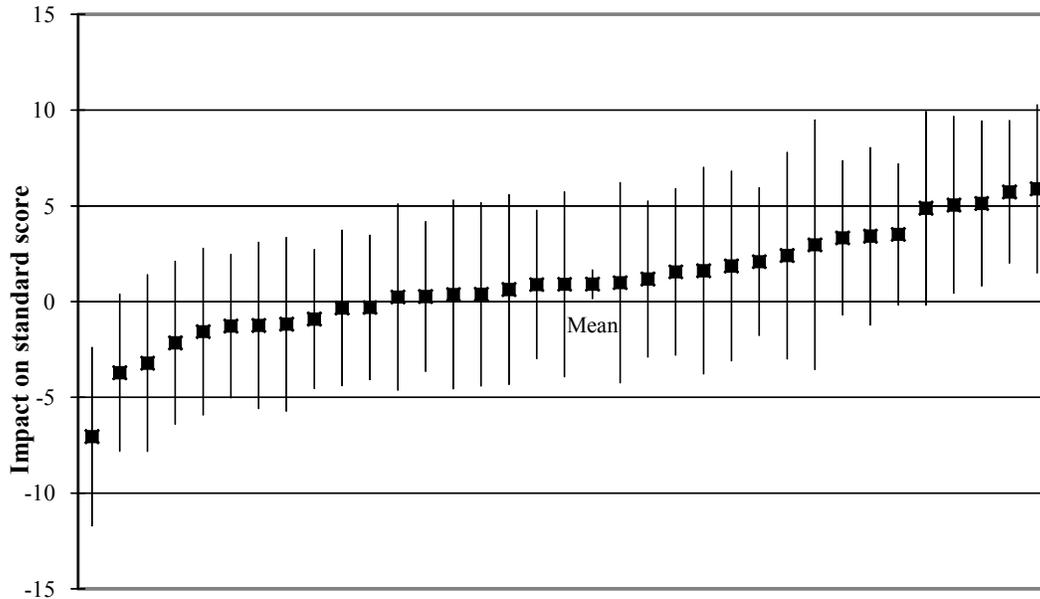
To examine variability in impacts across schools more systematically, a composite F-test was used to assess whether the school-level impacts on reading comprehension test scores

¹¹Estimated impacts are presented in numerical (ascending) order. See Appendix I for numeric values presented in Figure 5.2.

The Enhanced Reading Opportunities Study

Figure 5.2

Fixed-Effect Impact Estimates on Reading Comprehension, by School



SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment.

NOTES: The follow-up GRADE assessment was administered in the spring of 2006 near the end of students' ninth-grade year.

The fixed-effects impact estimates are the regression adjusted impacts of the interaction between school and treatment using ordinary least squares, controlling for blocking of random assignment by school and for random differences between the ERO and non-ERO groups in their baseline reading comprehension test scores and age at random assignment.

are statistically equivalent. This test accounts for estimation error in school-level impacts and provides an indication of the confidence one might have that there is variation in true impacts across the schools. The results show that the p-value for the F-test is 0.013, indicating that the school-to-school variation in impacts is statistically significant and, thus, is unlikely to have occurred by chance.¹²

¹²See Appendix I for the results of this F-test.

Impacts Associated with Implementation Fidelity and Duration During the First Year

First, the analysis examines impacts for groups of schools defined by whether the implementation of their ERO programs was classified as well aligned, moderately aligned, or poorly aligned with their respective program models, as defined in Chapter 3. This analysis provides insight into the hypothesis that ERO programs could produce stronger impacts if they are able to create classroom learning environments and to develop instructional strategies that were deemed to be relatively closely aligned with the specifications of the program that they were using. The top panel of Table 5.6 provides a summary of impact findings for the subgroups of schools defined by the implementation-fidelity categories that are discussed in Chapter 3.¹³ The first column shows the estimated impact on reading comprehension test scores, and the second column shows the estimated impact on vocabulary test scores. The far-right pairs of columns show estimated impacts on the three reading behavior measures. All impacts estimates are presented in effect size units.

The top panel of Table 5.6 indicates that, on average, the 16 schools whose ERO programs had reached a well-aligned level of implementation fidelity on both the classroom learning environment and the comprehension instruction dimensions of their models produced positive, but not statistically significant, impacts on reading comprehension test scores. A similar impact is exhibited in the third row for the 10 schools whose ERO programs were found to have poorly aligned implementation fidelity on at least one of the two dimensions. Statistically significant impacts were found for the eight schools whose ERO programs reached at least a moderately aligned level of fidelity on both dimensions but were not able to reach an adequate level on at least one dimension. In fact, the difference in impacts on reading comprehension test scores between the schools in the moderately aligned fidelity category and schools in the poorly aligned fidelity category is statistically significant. This result should be interpreted cautiously, however, because a composite test indicates that overall variation in impacts across the three fidelity subgroups is not statistically significant.

The top panel of Table 5.6 also provides a test of the linear relationship between impacts and a continuous indicator of overall implementation fidelity.¹⁴ The result presented in

¹³Appendix Tables I.2 through I.7 in Appendix I provide the outcome levels for the ERO and non-ERO groups, the estimated impacts, impact effect sizes, and p-values for the estimates presented in Table 5.6. The tables in Appendix I also show the differences in estimated impacts across school subgroups and p-values of these differences.

¹⁴For the purposes of this analysis, an indicator was calculated as the average of the fidelity rating for the classroom learning environment dimension and the fidelity rating for the comprehension instruction dimension. A value was calculated for each school ranging from one to three and rounded to the nearest tenth. The interaction between this indicator and the treatment indicator was added to the impact estimation model. The parameter estimate for this interaction term indicates whether the ERO program impact increased or decreased as a linear function of the fidelity indicator.

The Enhanced Reading Opportunities Study
Table 5.6
Impact Effect Sizes, by First-Year Implementation Issues

	Number of Schools	Reading Comprehension		Vocabulary		Amount of School-Related Reading		Amount of Non-School-Related Reading		Use of Reflective Reading Strategies	
		Impact	P-Value	Impact	P-Value	Impact	P-Value	Impact	P-Value	Impact	P-Value
		Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size	Effect Size
<u>Fidelity of program implementation</u>											
Well-aligned implementation	16	0.06	0.260	-0.05	0.404	0.04	0.466	0.06	0.282	-0.02	0.778
Moderately aligned implementation	8	0.22 *	0.005	0.17 *	0.027	0.16	0.057	0.13	0.120	-0.07	0.362
Poorly aligned implementation	10	0.02	0.797	0.03	0.655	-0.13	0.115	-0.07	0.345	0.06	0.433
Continuous fidelity measure	34	-0.04	0.498	-0.12	0.075	0.02	0.794	0.03	0.637	-0.06	0.400
<u>Duration of program implementation</u>											
More than 8.0 months	7	0.16 *	0.039	-0.09	0.258	0.05	0.579	0.05	0.593	-0.06	0.482
7.6 to 8.0 months	15	0.10	0.081	0.06	0.239	-0.01	0.922	0.02	0.745	0.02	0.675
7.5 month or fewer	12	0.02	0.712	0.05	0.487	0.04	0.570	0.07	0.339	-0.02	0.754
Continuous duration measure	34	0.07	0.351	-0.09	0.246	-0.01	0.896	-0.04	0.682	0.00	0.968
<u>First-year implementation issues</u>											
Moderately or well-aligned and longer duration	15	0.17 *	0.002	0.01	0.848	0.11	0.065	0.10	0.075	0.01	0.887
Poorly aligned or shorter duration	19	0.01	0.811	0.04	0.412	-0.07	0.250	-0.02	0.744	-0.02	0.695

SOURCE: MDRC calculations from the Enhanced Reading Opportunities Study follow-up GRADE assessment and follow-up student survey.

NOTES: Appendix I provides detailed information about each of the program implementation subgroup impact estimates, including outcome levels for the ERO and non-ERO groups, impact estimates, p-values, and differences in impacts among subgroups.

The estimated impact effect size is calculated as a proportion of the standard deviation of the non-ERO group average (reading comprehension = 10.458; reading vocabulary = 10.505; school-related reading = 43.867; non-school-related reading = 31.834; use of reading strategies = 0.670).

A two-tailed t-test was applied to the impact estimate. The statistical significance is indicated (*) when the p-value is less than or equal to 5 percent.

Table 5.6 indicates that the linear relationship between impacts and this overall fidelity indicator is not statistically significant.

Finally, the top panel of Table 5.6 indicates that, with the exception of vocabulary, impacts on other outcomes across the groups of sites are not statistically significant. Although the ERO programs in the moderately aligned fidelity category of schools produced a positive and statistically significant impact on vocabulary test scores, the difference in impacts across the subgroups is not statistically significant.

The analysis now turns to an examination of impacts for subgroups of schools defined by how long they were able to implement their ERO programs during the first year of the study. The length of program operation encompasses two first-year implementation challenges. First, delays in the start-up of the ERO programs meant that students randomly assigned to the ERO programs had already spent between three and 10 weeks enrolled in a regular elective class that they would have to leave in order to enroll in an ERO class. Rescheduling them into the ERO class was disruptive and required that they acclimate themselves to a new teacher and set of classroom routines. Second, the variation in the start-up delays meant that different amounts of time were available for teachers to cover the course curricula for the ERO programs and for ERO students to receive exposure to the ERO activities and materials that were planned by the developers.

The middle panel of Table 5.6 shows estimated impacts for three groups of sites: those that were able to operate for more than eight months, those that were able to operate for more than seven and a half months but less than eight months, and those that were able to operate for seven and a half months or less. The designation of these groups of schools — particularly those at either end of the distribution — reflects key differences in the potential interaction between implementation and program start-up, or duration. Schools that experienced start-up delays of six weeks or more — and that could operate for only seven and a half months or less — may reflect the most disruptive start-up for students assigned to the ERO classes and had the shortest amount of time to cover the ERO curricula. On the other hand, while none of the programs was able to operate for the full school year, by operating for more than eight months of the nine-month school year, schools had the opportunity to expose their ERO students to nearly 90 percent of the ERO learning strategies and activities offered by their programs. Schools in the middle group were able to operate their ERO programs between seven and half and eight months.

The middle panel of Table 5.6 shows that the estimated impacts on reading comprehension are positive and statistically significant (effect size = 0.16 standard deviation and p-value = 0.039) for schools that operated for the longest period of time during the school year. Although the differences in impacts across the three subgroups of sites are not statistically significant, the table indicates that estimated impacts are smaller for schools with shorter operating periods (effect sizes = 0.10 and 0.02 standard deviation). Table 5.6 also provides a test of the linear rela-

tionship between impacts and a continuous indicator of the number of months of ERO program implementation.¹⁵ The result presented in Table 5.6 indicates that the estimated linear relationship between impacts and month of program operation is not statistically significant, although the estimate itself is positive (effect size = 0.07 and p-value = 0.351).

Finally, the middle panel of Table 5.6 indicates that impacts on outcomes other than reading comprehension across the subgroups of sites based on ERO program implementation duration during the first year are not statistically significant.

Impacts Associated with a Combination of Challenges Relating to Implementation Fidelity and Program Duration

The analysis presented in this final section of the chapter attempts to shed light on the degree to which impacts may have been stronger in schools where the challenges associated with the combination of the implementation dimensions were less serious than in schools where implementation fidelity was poorly aligned with the program models or start-up was delayed by more than six weeks. As noted in Chapter 3, many of the challenges associated with implementation fidelity and delayed start-up that were present in the first year of the project have been addressed in the second year.

As discussed in Chapter 3, the implementation of the ERO programs in 10 of the high schools was classified as poorly aligned with their program models. Also, Chapter 4 discusses the fact that 12 of the high schools experienced delays of more than six weeks in the start of their programs as they struggled to recruit and enroll students in the ERO classes and the study sample. The implementation of the ERO programs in three of these 12 schools was also classified as poorly aligned with their program models. In all, therefore, the first-year implementation experiences of 19 of the 34 participating high schools can be seen as especially problematic, either because of inadequate implementation fidelity or because of particularly long delays in enrolling students in their ERO classes and the study sample.¹⁶

The bottom panel of Table 5.6 provides a summary of impacts for schools that were able both to reach at least a moderately aligned level of implementation fidelity and to operate for

¹⁵A value ranging from six months to eight and a half months was calculated for each school. The interaction between this indicator and the treatment indicator was added to the impact estimation model. The parameter estimate for this interaction term indicates whether the ERO program impact increased or decreased as a linear function of the length of time that the programs were in operation.

¹⁶This includes (1) seven high schools that experienced poorly aligned implementation, even though they were able to begin the classes within six weeks of the start of the school year; (2) nine high schools that experienced a start-up delay of more than six weeks, even though the implementation of their ERO programs ended up being classified as at least moderately aligned with their program models; and (3) three high schools that experienced both poorly aligned implementation and a start-up delay of more than six weeks.

more than seven and a half months during the first year of the study. The ERO programs in these 15 high schools reflect conditions that were closer to those intended by the design of the demonstration than in the remaining 19 high schools that did not meet one or both of these conditions.

The bottom panel of Table 5.6 shows first that the ERO programs produced positive and statistically significant impacts on reading comprehension test scores in the 15 schools where the ERO programs were classified as at least moderately aligned with the program model and began operation within six weeks of the start of the school year. The difference between the impacts on reading comprehension for these schools and for the remaining 19 schools is an effect size of 0.16 standard deviation. This difference in impacts is statistically significant and is consistent with the hypothesis that a combination of higher-fidelity implementation and a more timely start-up (longer duration) may contribute positively to stronger impacts on reading comprehension.

Conclusion

The early impact findings indicate that, overall, the literacy programs in the ERO study produced a statistically significant improvement in students' reading comprehension skills during the first year of implementation. The findings for the ERO programs that experienced a stronger start-up provide an indication of the effectiveness of the supplemental literacy programs under conditions more reflective of the intent of the ERO project. These conditions include implementation fidelity that was moderately aligned with the ERO program model and an operating period that was more than seven and a half months. In the schools where both of these conditions were in place, the ERO programs produced a larger impact on the reading comprehension skills of struggling adolescent readers.

Although the ERO programs produced some improvement in reading comprehension test scores, students in the ERO group continued to lag behind the average ninth-grade student nationally. The 90.1 average standard score achieved by students in the ERO group at the end of their ninth-grade year corresponds, approximately, to the 6.1 grade equivalent and the 25th percentile nationally. Even when schools that experienced the most significant challenges with the first-year implementation are excluded, the more substantial impact on reading comprehension test scores for the remaining schools still left many students well below grade level. In fact, almost 90 percent of the students in both the ERO and the non-ERO group were still reading below grade level at the end of their ninth-grade year, and 76 percent of the students in the ERO group were two or more years below grade level and, thus, would still be eligible for the ERO programs, as specified by the criteria used for this project.

The early impact findings discussed in this report do not represent conclusive evidence about the efficacy or effectiveness of the supplemental literacy interventions being tested. Recognizing the need for the participating schools and teachers to gain more experience with the

programs, the U.S. Department of Education built into the design of the ERO project a second year of implementation and a second cohort of ninth-grade students for the study sample. The next report from the ERO study will provide evidence on the impact of the supplemental literacy programs during this second year of implementation. A critical goal of the second year of the study is for the participating schools and teachers to address the start-up challenges that arose in the first year and to apply their experiences from the first year and subsequent additional training.

As of this writing, the ERO study has begun to examine implementation data from the second year of the study. Twenty-seven of the 34 teachers who taught the ERO classes in the first year of the study returned for the second year. These teachers and the seven replacement teachers participated in a summer training institute and continued to learn more about how to use the instructional strategies that lie at the heart of the two interventions. All these teachers remained with their ERO programs throughout the second year. A second cohort of ninth-grade students was identified for the 2006-2007 school year. Across the 34 schools, ERO classes began within an average of approximately two weeks of the start of the school year and, at 18 of the schools, began on the first day of school.

The ultimate goal of the two ERO programs is to improve students' academic performance during high school and to keep them on course toward graduation. With this in mind, subsequent reports from the evaluation will examine the impact of the programs on student performance in their core academic classes, their grade-to-grade promotion rates, and their performance on high-stakes tests required by their states. The final report from the project will present impacts on these outcomes through the eleventh grade for students in the study's first cohort and through the tenth grade for students in the second cohort.